“Horses in the Ether”: Towards a Social Semantic Knowledge Sharing and Collaborative Initiative for the Thoroughbred Bloodstock Industry

Michael Hartnett

A dissertation submitted in partial fulfilment of the requirements of
Dublin Institute of Technology for the degree of
M.Sc. in Computing (Knowledge Management)

March 2011

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Knowledge Management), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: _____

Date: ***21st March 2011***

ABSTRACT

Since its establishment, over 250 years ago, the Thoroughbred Bloodstock Industry has accumulated vast stores of information and knowledge about Thoroughbred Horses, their pedigrees, races and results. The ability to share this knowledge in a timely and efficient fashion is of paramount importance to the continuing success of the industry. Combined with this, the tacit and explicit knowledge that is retained by the members of the Thoroughbred Bloodstock Community indicates a vast wealth of unexploited knowledge that can be leveraged for the benefit of all members of the community. This dormant reserve of knowledge has been recognised by a number of high profile Industry Organisations who are actively involved in promoting the growth of a Learning Thoroughbred Industry, through dissemination of some of the available knowledge and experience.

The objective of this research is to investigate the concepts associated with Knowledge Management and its implementation in the context of web-based industry or Virtual Community of Practice. An analysis of the current technologies utilised for Knowledge Sharing within the Thoroughbred Bloodstock Industry will be investigated, e.g. RSS feeds, blogs, etc., with the intention of determining if an appropriate semantic ontology has been used to categorise and organise the knowledge for effective sharing. The intention of this analysis will be to explore how Web 2.0 technologies can be effectively leveraged, in conjunction with such Semantic ontologies as SIOC and FOAF, as a Knowledge Provider for the Thoroughbred Bloodstock Community.

An evaluation of the functionality of a proof-of-concept Knowledge Portal combined with the results obtained in this research are indicative of the feasibility of a Social Semantic Knowledge Sharing and Collaborative initiative by means of community-driven folksonomy and ontology construction.

Keywords: *knowledge, knowledge management, knowledge sharing, tacit knowledge, explicit knowledge, knowledge sharing behaviour, collaboration, communities of practice, web 2.0, social software, weblog, RSS, portal, tagging, tag cloud, text cloud, folksonomy, taxonomy, ontology, SIOC, FOAF, bootstrapping.*

ACKNOWLEDGEMENTS

I would like to thank the following people for their much appreciated support, encouragement and help throughout the completion of this research.

First thanks to my dissertation supervisor Ronan Bradley for his guidance and advice throughout this research.

I would also like to thank, most sincerely, Damian Gordon, for the unique knowledge he provided through the modules he taught, for the support and advice he offered and for the devotion he showed to this research……."***Dude***"

I would also like to thank all of those people who participated in and completed the survey used for this research.

Very special thanks to my enduring wife, Helen Hartnett, for all her help, love, support, encouragement and supply of coffee during this research. Thanks also to my wonderful children Saraid, Tomás and Meadbh who have patiently waited for their Daddy to finish his college work.

Finally, I would like to thank all my family and friends for helping and supporting me throughout the year.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

1.1 Background

The origins of the Horse Industry can be traced as far back as 4,500 B.C., when the prehistoric nomadic tribesmen of Central Asia first domesticated the horse. During the 12th Century, English Knights returning from the Crusades, brought with them Arab Horses, which were the origins of the Thoroughbred breed when the Arab Stallions were bred with English mares during the 17th and 18th Centuries. The Jockey Club was created in 1750 to govern the emerging sport of Thoroughbred Horse Racing and also set up the General Stud Book to record the pedigree of all racing Thoroughbreds, now maintained by Weatherbys. An American Stud Book was first established in 1868 and by 1894 the American Jockey Club was set up (Horseracing.com, 2011).

Since then the Thoroughbred Bloodstock Industry has expanded and evolved into the global sport, hobby, passion and business which is currently recognised worldwide today. It can be seen from this brief history of the industry that the management of over 250 years-worth of knowledge of Thoroughbred Horses, their Pedigrees, Races and Results, and the ability to share this knowledge is of paramount importance to the continuing success of the industry.

The ability to share knowledge within the domain of the Thoroughbred Horse Industry has become common practice, by virtue of the accessibility of the World Wide Web and the small investment required to set up websites, knowledge sharing portals and blogs. While this accessibility has resulted in a wealth of knowledge being shared within the Bloodstock Community, it has also resulted in a diversity of knowledge, as multiple sources offer multiple versions of the same information, albeit in different formats. As a result the domain has become overwhelmed by the extensive amount of the same knowledge, with no way to effectively consume it.

The numerous disparate sources of information contributing to Knowledge Sharing within the Bloodstock Community do not openly contribute to a collaborative approach, as each expert / individual provides their own conceptualised version of

industry knowledge. This approach does not embrace the best practices for Knowledge Sharing, instead it directly conflicts with the concepts and proves inefficient in practice.

At present various sources of knowledge exist, e.g. blogs, news feeds, calendars, PDF and XML newsletters, but these sources need to be better leveraged to provide a common source of knowledge. The knowledge provided by these sources is highly structural due to the hierarchical nature of the areas within the industry. This knowledge should, therefore, be highly queryable, searchable and accessible, if presented correctly, see Figure 1, Table 1 and Table 2.

The intention of this dissertation will be to explore which Web 2.0 tools are being used by the Thoroughbred Bloodstock Community and how they can be effectively leveraged as a Knowledge Provider for the Thoroughbred Bloodstock Industry. Additionally, an examination of current web semantics and RSS ontologies will be undertaken to suggest how they can be extended to contribute to improving the efficiency and effectiveness of sharing this knowledge.

1.2 Knowledge Sharing

Over the last number of decades the interest in organisational learning and so-called "learning organisations" has grown tremendously. The capacity of an organisation to learn effectively, or to develop what Peter Senge calls "a learning capability" (Senge, 1999), rather than just emphasising dispersal of information, is increasingly seen as necessary in order to effectively share knowledge.

The Merriam-Webster Dictionary defines knowledge as "the body of truth, information and principles acquired by humankind" (Merriam-Webster A, 2010). The culture of Knowledge Sharing can therefore be defined as "the integrated pattern of human behaviour that includes thought, speech, action, and artefacts and depends on man's capacity for learning and transmitting knowledge to succeeding generations" (Merriam-Webster B, 2010). This definition of a Knowledge Sharing Culture is important as increasing the capacity for learning and sharing knowledge is one of the principal aims of Knowledge Management.

The concepts of Knowledge Sharing and Collaboration are fundamentally supported by the formation and evolution of Communities of Practice. Communities of Practice are formed by people who engage in a process of collective learning in a shared domain. Cognitive anthropologists Jean Lave and Etienne Wenger (1991) defined a Community of Practice (CoP) as "a group of people who share a … common interest in a particular domain or area … with the goal of gaining knowledge related to their field" (Lave & Wenger, 1991). According to Seely Brown & Duguid (1991) members of a Community of Practice are thought to be more efficient and effective conduits of knowledge (Seely Brown & Duguid 1991), while Ardichvilli, Page & Wentling (2003) support this indicating that individuals that are members of CoPs report increased communication with other individuals, helping to generate new knowledge (Ardichvilli, Page & Wentling 2003).

Communicating with others in a community of practice involves creating "*social presence"*, which Tu (2002) believes affects how likely an individual is of participating in a CoP, especially in online environments (Tu, 2002). CoPs can exist in online environments as Virtual CoPs (VCoP) within discussion boards and newsgroups, or in real life as Real CoPs through face-to-face social interaction e.g., at work, on a factory floor, or elsewhere in the environment. Wenger, White & Smith (2009) argue that virtual communities change the way we think of community and that technology stewardship is a key element of a VCoP (Wenger, White & Smith, 2009). While Web 2.0 technologies and the online environment facilitate the development of VCoPs, Kimble, Hildreth & Wright (2001) have found that facilitation of participation is central to the evolution and success of the VCoP (Kimble, Hildreth & Wright, 2001). The facilitation of participation is more difficult in a VCoP as real-time face-to-face interaction is not present as opposed to a Real CoP where personal interaction is vital.

The Thoroughbred Bloodstock Industry can be regarded as a Community of Practice formed by numerous groups of business people, experts and individuals who share an interest and/or knowledge in the area of Thoroughbred Horses. Although such a Community of Practice exists, very often the Knowledge Shared within the community is of a disparate and diverse nature. Much of this is directly related to the expertise of

the individual sharing the knowledge, individuals ranging from Amateur Race Goers and Breeders, Industry Pundits, Breeders, Trainers, Jockeys and Racing Organisations.

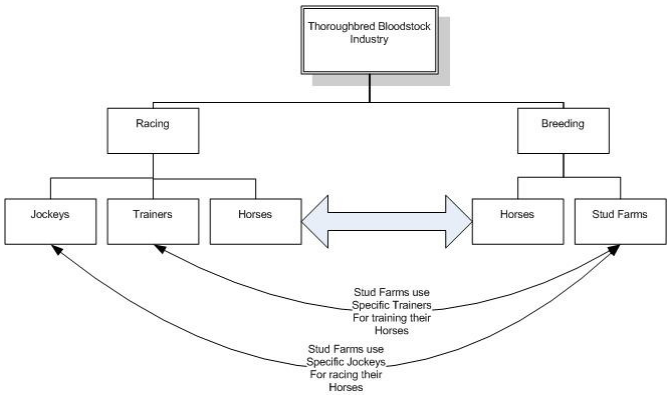


Figure 1 Hierarchical nature of Thoroughbred Bloodstock Industry

The disparate and diverse Knowledge Sharing within the online Bloodstock Community will be the main emphasis of this dissertation, with particular interest placed on numerous aspects of the knowledge dispersal, including domain expert blogging, e.g. Horse Racing and Breeding Blog and professional RSS feeds.

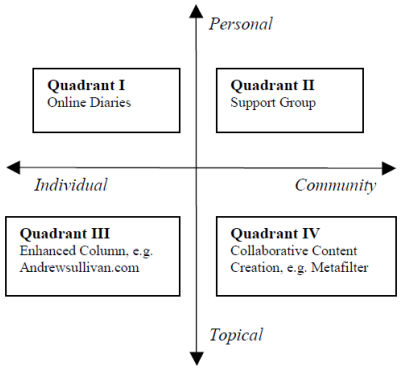


Figure 2 Types of Blogs

A number of studies on the different types of blogs have been produced with Krishnamurthy (2002) proposing a classification of blogs into four basic types according to two dimensions: personal vs. topical, and individual vs. community (Krishnamurthy, 2002). This project will primarily examine blogs that fall into Quadrants II and IV.

Summerhill Blog 	URL: http://blog.summerhill.co.za/ Year of Creation: 2006 Interesting, Informative & Knowledgeable - Maintained by Summerhill, one of South Africa's largest and most well-known stud farms, initially as a marketing and promotional tool, but has evolved into a comprehensive repository of knowledge concerning the South African Thoroughbred Industry.

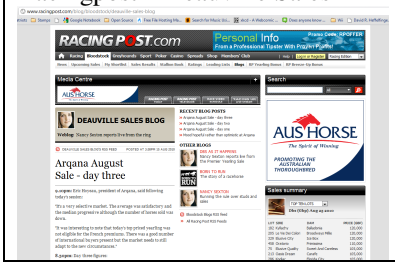
Racingpost – Born to Run 	URL: http://www.racingpost.com/blog/bloodstock/born-to-run/ Year of Creation: Unknown Interesting - "The story of a racehorse" – Provides insightful information and knowledge about all aspects of breeding of racehorses.
Calvados Bloodstock 	URL: http://www.calvadosbloodstock.com/ Year of Creation: 2010 Knowledgeable - Maintained by a graduate of the Irish National Stud's Thoroughbred Breeding and Management Course, mainly related to information about Thoroughbred Pedigree Research.
Racingpost – Deauville Sales 	URL: http://www.racingpost.com/blog/bloodstock/deauville-sales-blog Year of Creation: Unknown Informative - A seasonally available blog site, relating information about Thoroughbred Horse sales in Deauville, France.

Table 1 Sample Weblogs

1.3 Research Problem

The motivation for the project has arisen from 11 years working experience in the Thoroughbred Bloodstock Industry and subsequent indepth knowledge and domain expertise. During that time, through various projects undertaken to assist clients worldwide consume some of the rapidly expanding varieties and sources of information available, it became clear that a lot of the same knowledge was being captured in multiple diverse formats and from multiple disparate sources. Some of these multiple formats and sources used include:

- XML racing feeds – supplied by RISA (Racing Information Services Australia)
- XML based pedigree and breeding records – supplied by TJCIS (The Jockey Club Information Systems – America)
- Text based racing feeds – supplied by The Racing Post (UK) & The Emirates Racing Authority (Dubai)
- RSS feeds from Blog and News sites – Racingpost, Summerhill Blog, etc.

- PDF newsletters from Thoroughbred Breeding Organisations – EBN (European Breeders News)

Additional development of web based Content Management Systems and Information Portals for a number of clients has provided a background for the research and development of this project.

It is proposed that an open source knowledge portal be used to consume aggregated RSS Feeds delivered by varying sources of knowledge within the online Thoroughbred Bloodstock Community. The content of each feed will be consumed and stored in a database in order that a semantic analysis can be carried out. The expected results will be based on the analysis of existing online resources as well as an experimental portal, developed specifically for the purpose of this investigation. Online surveys, observations and relevant literature will also be used.

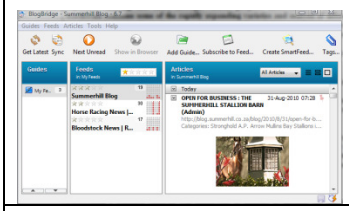
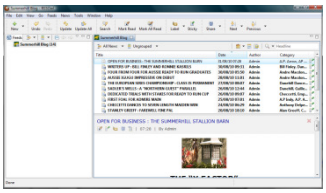
BlogBridge 	URL: http://www.blogbridge.com/ Year of creation: 2007 (Stable Release) An Open Source Java Based RSS feed aggregator that can combine multiple RSS and blog feeds into a single searchable resource.
iGoogle 	URL: http://www.google.com/ig Year of creation: 2005 A web based information portal and dashboard that can incorporate multiple RSS feeds and news resources.
RSSOwl 	URL: http://www.rssowl.org/ Year of creation: 2003 Another Open Source RSS feed aggregator which includes advanced grouping, searching, labelling and exporting of feed articles.

Table 2 Free & Open Source RSS Aggregators

Performing a semantic analysis will attempt to identify if, initially the appropriate RSS specific semantic information has been supplied as part of the feed articles. This will include Gap Analysis to determine what is included and what has not been included.

An investigation will be to undertake an analysis of existing RSS feeds and tagging methods and to propose a common ontology or extension to an existing ontology to

facilitate more comprehensive and effective knowledge sharing within the Thoroughbred Bloodstock Domain.

Further analysis for existing semantic ontologies will be undertaken to identify any existing Thoroughbred Bloodstock related, or domain-specific, ontologies. If existing ontologies can be identified, they will be applied to the RSS feed data to try to establish their applicability and whether the RSS feeds have been semantically developed. Should an appropriate ontology not exist then a proposed extension or new ontology will be presented as part of the project. This will represent an Open Source Semantic Knowledge Sharing Initiative that will embrace the best practices for Knowledge Sharing in the online Bloodstock Community.

1.4 Research Objectives

The following research objectives are expected to be achieved through this research and contribute to the overall outcome:

1. Examine current Learning and Knowledge Sharing methods in the Thoroughbred Bloodstock community and identify deficiencies and areas for improvement.
2. Investigate how the web can be used as an integral source of Knowledge in the Thoroughbred Bloodstock Community.
3. Use Web 2.0 tools to retrieve and store information.
4. Investigate and propose Industry specific Semantic Ontology and Knowledge Sharing Specification.
5. Propose an environment to facilitate an online Community of Practice where information can be shared.
6. Evaluate appropriateness of proposed ontology and industry perception of a Knowledge Sharing initiative, through the design and evaluation of a proof-of-concept prototype.

1.4.1 Research Methodology

This research will begin using an anthropological approach, exploring what sources of information exist that are relevant to the Thoroughbred Bloodstock Industry, this will

be a wide-ranging search that will look at a variety of sources, including blogs, news sites, journals and homesites of important organisations. The purpose of this exploration will be to examine existing sources of information, to identify both the overlaps and the gaps between sources.

An important part of such a project is to define all the research methodologies used during the research, design and evaluation phases of the project.

- *Qualitative Research* - This will form the basis of a critical review of literature available in the domain of Knowledge Management to provide an insight into its use in the Thoroughbred Bloodstock Domain. This includes the examination of Knowledge Management literature from the books, journals, online resources, white papers and resources available on the internet.

- *Quantitative Research* - The collection of data through the use of questionnaires or surveys will be examined. A survey will be designed to gather information about the potential for knowledge sharing in the Thoroughbred Bloodstock Industry.

- *Experimentation* – An experiment will be conducted, based on the Qualitative and Quantitative Research which will be used to examine and propose the potential for a Social Semantic Knowledge Sharing initiative for the Thoroughbred Bloodstock Industry. Further experimentation will be undertaken by virtue of the design and development of a proof-of-concept prototype to model this Knowledge Sharing initiative.

- *Interviews* - The use of interviews will provide the author with an insight into the reactions and comments of Industry Experts in relation to the topics being researched and presented.

- *Observation and expert insights* - Observation and the insights of experts on this topic will be used to gather information that is related to the evaluation of the outcomes of the experimentation phase of the project.

All these research methodologies will be used to complete this research and achieve the specified research objectives.

1.4.2 Scope and Limitations

The scope of this dissertation is to examine the current knowledge management practices which occur in Thoroughbred Bloodstock Industry and to determine their effectiveness as mechanisms for knowledge sharing and collaboration. This will take the form of extensive literature review and 11 years working experience and industry observation, enabling the author to ultimately recommend a solution.

The author acknowledges that the research to be carried out will not be exhaustive, but will serve to identify the most relevant resources and documentation available, to be able to gain insight into the identified problem and associated difficulties. Though a great deal of Knowledge Management research and case study literature exists, there is very little which deals with the subject of the Thoroughbred Bloodstock Industry. Though, this can be considered to be a limitation to fulfilling the aim of the project, the author has decided it will be necessary to explore and review existing Knowledge Management literature and apply this in the context of this research.

The purpose of this research is not an attempt at creating a database of Thoroughbred terms. Nor is it attempting to create a complete Ontology for the Thoroughbred Bloodstock Industry, but rather strives to highlight the fact that one does not exist. Due to the specialist nature of the industry, and its multitude of domain experts, the research indicates that the knowledge of these experts could be elicited to develop a working ontology through the concepts of Communal Tagging and Folksonomy Formalisation.

The conceptual portal that will be proposed will not be designed as a one-way vessel of Knowledge. In order that a successful initiative can be proposed, its potential users will be required to contribute and update the knowledge within the portal. This community contribution will be essential for the nurturing of Knowledge Sharing through a collaborative industry approach.

Similarly the portal will not be a working model but a proof-of-concept which, the author hopes, will encapsulate and model the concepts of Knowledge Sharing and Collaboration in a fashion that will be informative and practicable.

1.5 Organisation of The Dissertation

Chapter 2 will start by introducing the Thoroughbred Industry in terms of its members, their knowledge and their collective ability to learn using various methods of work activities. It will also examine the foundations of Knowledge Management in terms of a Knowledge Industry discussing how Communities of Practice can be used for Knowledge Sharing. Finally, the current paradigm shift in Knowledge Sharing, from the traditional centralised repository to a more conversational collaborative intelligence, gives rise to a presentation of a Virtual Community of Practice which uses Web 2.0 and Social Software as mediums for Knowledge Sharing.

Chapter 3 will examine how the use of Web 2.0 technologies provide an ideal background of research promoting the use of Weblogs and their RSS technology as a Knowledge Management and Sharing Tool

Chapter 4 will examine the use of RSS as a Knowledge Sharing tool viewed as the driving component of Portals on the Web. This chapter also seeks to inform the reader of the progressive nature of categorisation and classification techniques used, from the informal (Social) to the formal (Semantic), through the use of Tags, Folksonomies, Taxonomies and domain-specific ontologies. The chapter will conclude with a brief critique of the use of Portals in the Thoroughbred Bloodstock Industry.

Chapter 5 describes the research undertaken in order to gain a better insight into the use of blog technologies in the Thoroughbred Industry, as a key enabler for the purposes of knowledge sharing.

Chapter 6 will review the current organisation of knowledge within the Thoroughbred Bloodstock Industry investigating its adequacy for the conveyance of organised knowledge. An examination of the content of syndication feeds will assess their suitability for the development of an industry folksonomy. This chapter will also

produce and review an emerging industry folksonomy and how it may contribute to development of a basic industry specific terminological ontology. An examination of the extension of the ontology using Semantic Web frameworks, such as SIOC (Socially Interlinked On-line Communities) and FOAF (Friend of a Friend) will be discussed in terms of a Semantic Bootstrapping process for the Industry.

Chapter 7 will present, discuss and evaluate a proof-of-concept prototype for a potential Thoroughbred Industry Knowledge Portal. A design and usability evaluation will be conducted based on the two main concepts being modelled; Social Networking and Communal Tagging. Feedback from the evaluation audience will be presented and analysed.

Chapter 8 presents the conclusion of the project, highlighting its contributions to the body of knowledge, evaluations and limitations, and proposes potential areas for future work.

1.6 Dissertation Objectives

The deliverables will include a project dissertation containing the relevant literature reviews in the area of The Thoroughbred Bloodstock Industry and its Knowledge Management practices, Web 2.0 and online Knowledge Sharing among Communities of Practice.

All survey and questionnaire literature will be delivered, with relevant analysis to determine levels of knowledge sharing. Additional analysis will be presented in the form of gap analysis, based on the semantic nature of the feed details, to determine what elements are missing, irrelevant and those that need to be added.

A proof-of-concept web based portal will be presented as part of the project which will be used model the conceptualised Knowledge Sharing mechanisms proposed after analysis of RSS feeds and their content

A Glossary of all relevant industry terms will be supplied as part of the project submission, derived from an emergent Industry Folksonomy.

2 KNOWLEDGE AND LEARNING IN THE THOROUGHBRED INDUSTRY

2.1 Introduction

Learning and Knowledge Sharing can best be described as an act of social interaction, and can be seen as aiding in the development of both tacit and explicit knowledge. This is evident in the Thoroughbred Industry where the daily work activities of professionals can be viewed as a medium for learning and knowledge sharing. However, the scope of knowledge sharing is limited to the contact and interaction, within the workplace, and therefore provision must be made in order to share the wealth of knowledge in the industry on a global scale. Traditional Knowledge Management activities are generally studied in the context of individual organisations. Knowledge Sharing is implemented through use of central repositories, which can be accessed and added to by the knowledge workers within the organisation. However, in the context of a Knowledge Industry, no single centralised database or technology system can fully capture and distribute all the knowledge contained within an industry. This is particularly true of the Thoroughbred Industry, where the global nature of the industry and multi-dimensional aspect of its work domains, means that a diverse set of knowledge is created which can be of interest to all domains in the industry. The Thoroughbred Industry must therefore adapt and evolve in order to fully leverage new technologies and innovations, for the provision of an industry-wide Knowledge Sharing culture, by endeavouring to become a Knowledge Industry.

This chapter will look at the Thoroughbred Industry through the prism of knowledge management, and will consider issues such as Communities of Practice (Real and Virtual), Knowledge Sharing, and the differences between the traditional organisational-based view of knowledge management and how this differs in the Thoroughbred Industry.

2.2 The Thoroughbred Bloodstock Industry

'... an industry that puts Ireland on the map right around the world.... we are very, very proud of it.... At the end of the day it's a vision for our country to be the best and to keep on being the best in an industry that, by your efforts, you have made Ireland's name soar throughout the world and long may that continue.'

President MaryMcAleese, September 2009.

Gross Value of Industry to the economy 2008 (€)	
Training fees 2008 ‡	289,810,761
Exports *	216,000,000
Festivals *	260,600,000
Irish coverings ‡	124,181,050
Irish domestic public sales	57,207,780
Stock keep fees	225,300,240
Total	1,173,099,831

Source : Calculations based on ITBA surveys ‡, ITBA calculation based on HRI data and confidential survey of retail bookmakers †, CSO data *

Figure 3 Gross Value of the Thoroughbred Industry to the Irish Economy in 2008

According to Irish Economist Alan Dukes (2009), "*Ireland's dominance in the industry is reflected in the involvement of Irish people on breeding and racing farms and in ancillary services throughout the world. Irish racing and breeding managers, veterinarians, farriers and other professionals are at the forefront of the industry on every continent, applying their skills and expertise.*"

This statement taken in a global context portrays the Thoroughbred Bloodstock Industry as a multi-faceted and multi-dimensional industry, characterised by its high levels of diversity in both expertise and knowledge. The industry has a strong global presence, in which distance contributes largely to the difficulty of developing a common international knowledge sharing mechanism that is within the acceptable limits of all industry members. Even in these current times of Economic Recession, the Industry has remained unified by virtue of its key asset, the Horse and Knowledge thereof. The Horse is used for both recreational and commercial business purposes and the knowledge of the industry members remains critical to the industry's on-going development and growth.

However, due to the historical background of industry fragmentation, tension and competition, the Knowledge upon which the industry relies could very well contribute

to its collapse if a suitable knowledge sharing mechanism is not introduced. The dissemination of knowledge, within the Thoroughbred Industry, must become the responsibility of the industry members, through improved and more efficient knowledge sharing initiatives. The introduction of a combined and community based knowledge sharing initiative could be seen to guarantee the competitive performance and future of all sectors of the Thoroughbred Industry.

2.2.1 The Learning Industry

Chris Morrison (2000) characterises the approach to knowledge management and knowledge sharing in the Thoroughbred Industry, as both conservative and adversarial, with many of its members originating from backgrounds of limited education. Working long hours and with restricted opportunities for conventional education, industry members have learnt using a bottom up approach in the "school of hard knocks" gaining knowledge and experience through apprenticeships.

The Thoroughbred Industry is built on the master-servant approach to apprenticeship training, demonstrating the method of learning in the industry. Morrison indicates that great improvements have been made in numerous sectors of the industry with regard to more accessible education and learning, but areas of resistance still remain that are reluctant to change.

Another inhibitor to learning opportunities acknowledged within the Thoroughbred Industry is that of the development of siblings who have sustained family heritage by continuing family legacies, businesses and traditions. Such examples are commonplace in the Irish Thoroughbred Industry, examples include the O'Brien family which has produced one of Ireland's most successful Father-Son horse training legacies with Vincent O'Brien and his sons David and Charles. This type of development is viewed by many to be an elite element within the Thoroughbred Industry that limits the learning prospects for members from families not connected with the industry.

Members of the Thoroughbred Industry have a multiplicity of learning requirements, primarily determined by an individual's requisite to learn and acquire new knowledge. In order to facilitate these learning requirements industry knowledge must be readily

available in a timely and flexible manner, to meet the demands of the learners. However, due to the rigorous nature of jobs and working hours within the Thoroughbred Industry, ordinary learning methodologies, such as classroom teaching, do not always support these requirements. Thus, it is important to find complementary methods of education that will match the requirements of the Thoroughbred Industry learner.

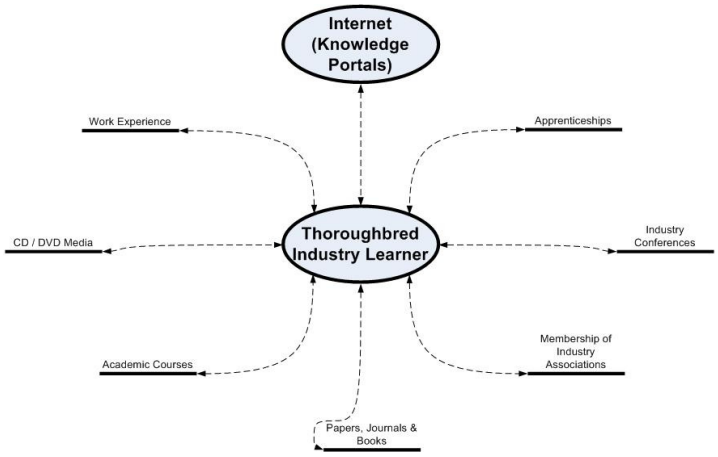


Figure 4 Insight into learning sources used in the Thoroughbred Industry

Some experts now believe that learning has moved beyond the realm of the organisation and are investigating the possibility of Learning Sectors and Industries. Antonacopoulou et al. (2006) indicates that the wider socio-cultural context at both organisational and industry levels has been a determining factor in the attitudes industry experts and professionals develop towards education, training, learning and development within an Industry. In the context of this dissertation this can be seen also as a determining factor in the attitudes industry experts and professionals display towards Industry-wide Knowledge Sharing.

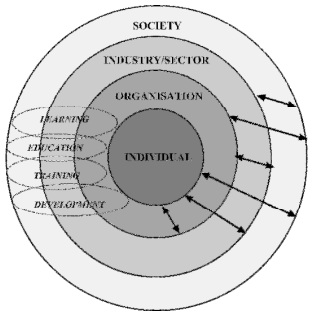


Figure 5 Conceptualisation of the relationships between education, training, learning and development, emanating from the interaction between individual, organisational, industrial and societal learning (Antonacopoulou et al., 2006)

One of the major elements of 'A learning Industry' is the realisation of industry experts that the shared effort of individuals is more powerful than the ideas of a single expert. Thus the learning industry must actively develop an environment that encourages its members to make a contribution to the body of knowledge that is beneficial to the industry-wide community.

Morrison (2000) acknowledges that an emphasis is being placed on the development of individuals who can accept change and adapt in a developing learning industry. He further indicates that positive attitudes towards training and development have resulted in the Thoroughbred Industry embracing the concept of a learning industry, through businesses and service providers who explicitly acknowledge the value of workplace learning and create an environment that enables knowledgeable experts to become actively involved in the learning process.

Examples of Industry Learning Enablers

A number of high profile Industry Organisations are actively involved in establishing and promoting the growth of a Learning Thoroughbred Industry. The following is an example of some of the leading industry organisations making an active contribution towards establishing a Learning Thoroughbred Industry.

The Irish National Stud, introduced its first Thoroughbred Breeding course in 1971 and it remains the best-known equine training programme to this day. The aim of the course is to educate young people for a career in the breeding industry, although graduates have also been prominent in racehorse training, bloodstock sales and insurance and the media. (source - http://www.irish-national-stud.ie/stud/breeding-course/)

Founded in 2003 by Sheikh Mohammed, Darley Flying Start has rapidly developed from a vision into a sophisticated management training programme "*to advance standards and knowledge in the management of thoroughbred horses*". (source - http://www.darleyflyingstart.com/dfs_background.shtml)

The English National Stud provides a variety of courses, many of which are free, aimed at those wishing to enter into a career in the Thoroughbred breeding industry. The courses combine practical, "hands on" training with theoretical knowledge, and the opportunity to gain recognised Thoroughbred industry qualifications. (source - http://www.nationalstud.co.uk/education.asp)

Each of these learning programmes provides education and learning advancements for the members of the International Thoroughbred Community. Enhancing the development of a Learning Thoroughbred Industry is the core edict of each of these organisations, through use of a multitude of learning experiences including global placements and interactive support. These organisations can thus be considered as enablers of the Thoroughbred Learning Industry which, through further development, are instrumental in the development of a Knowledge Industry.

2.3 Becoming a Knowledge Industry

Over time people accumulate the knowledge and skills they require that enables them to work more effectively and efficiently. Morrison (2000) proposes that the authenticity of information and knowledge related to learned practices, in the Thoroughbred Industry, is "hard to trace and appears to have been perpetuated through folklore and storytelling".

For centuries people have gained knowledge through experience, from other industry experts and through their mistakes when working within an industry. Examples of this manner of learning have been discussed in Section 2.3.1, i.e. apprenticeships, sibling development, etc. Field and Ford (1995) express this style of knowledge acquisition as disorganised, "For the people involved, learning is hit and miss".

Morrison (2000) further contends that a similar difficulty with knowledge today is still being experienced within the Thoroughbred Industry. He describes examples of respected industry experts who "*do not check the origin of their information*" and discusses the concept of the FOG factor presented by Straker (1997) as a means of checking the authenticity of the information and knowledge. Determining the authenticity of knowledge requires that:

- Facts can be proved to be true.
- Opinions are what people believe to be true.
- Guesses are acknowledged as wild ideas.

With this concept in mind it is important to develop a process that enables knowledge to be filtered and appropriately organised within an industry. Such an approach to the management of knowledge can be interpreted as a significant move towards becoming a Knowledge Industry.

2.3.1 Types of Knowledge

According to the Financial Times Lexicon, a Knowledge industry is an *"industry where success depends on obtaining, managing, and using knowledge in a particular area"*. (Financial Times Lexicon, 2011) In order to develop the concept of a Knowledge Industry, it is first imperative that the types of knowledge available in the Industry be identified and defined. Knowledge in the context of Industry is a diffusion of collective information, identified in different forms and by various methods. Nonaka and Takeuchi (1995) proposed the terms tacit knowledge and explicit knowledge as the two main types of human knowledge.

Tacit knowledge is the knowledge contained within an individual's heads defined as "*A form of knowledge that is highly personal and context specific and deeply rooted in individual experiences, ideas, values and emotions*". Tacit knowledge can be viewed as the most discernable form of knowledge evident within the Thoroughbred Industry. Horse Trainers, Breeders, Jockeys, etc. all acquire huge amounts of tacit knowledge through years of experience of dealing with Horses. Too often this knowledge is retained or shared in only limited capacity by the individuals in order that they can maintain an advantage within the industry.

Tacit knowledge which is gained through experience is also often difficult to articulate, which may also account for the often limited formalisation and communication of an individual's knowledge. Usually, people make use of personal originality to explain

and express their facts, opinions and guesses to communicate personal tacit knowledge to others (Steward 1998).

Alternatively, explicit knowledge can be simply defined as captured knowledge, which Polyani (1983) explains is codified, more formal and easier to transmit. Explicit knowledge is usually stored in books, journals, internal or external databases, different forms of digital knowledge bases, and Communities of Practice or online in virtual Communities of Practice.

The Thoroughbred Industry's very foundation is defined upon an accumulation of explicit knowledge stored and maintained by leading global industry organisations. While access to this explicit knowledge is available it is often limiting to individuals due to prohibitive costs for retrieval and transfer of ownership, even for information relating to "*One's own Horses*".

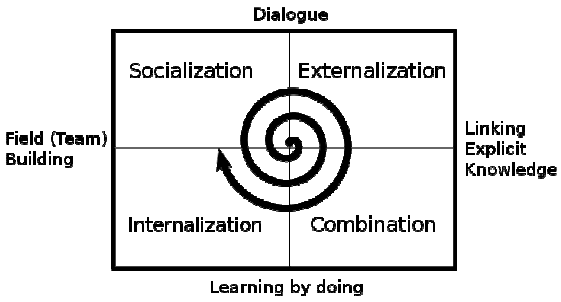


Figure 6 Knowledge Industry Framework, based on the Knowledge Spiral proposed by Nonaka and Takeuchi (1995)

Nonaka and Takeuchi (1995) propose a knowledge spiral (shown above as a Knowledge Industry Framework). Knowledge within the spiral, or framework, moves through a number of stages:

- Socialisation - Tacit knowledge is transferred between individuals by way of verbal transferral and non-verbal observation and practice as well as on-the-job training and mentoring, e.g. apprenticeships.
- Externalisation - Tacit knowledge is stored and codified in a formal manner making the knowledge available to all. In the context of this research making tacit knowledge explicit requires that a shared or common meaning be created

for concepts. (This concept will be further investigated in Chapter 4, through development of an Industry Folksonomy.)

- Combination - Explicit knowledge is combined with additional contextually compatible explicit knowledge and new knowledge is created
- Internalisation - Explicit knowledge is consumed by an individual and incorporated as part of their own tacit knowledge base.

This is an active learning framework where industry members share tacit knowledge, incorporate explicit knowledge and develop new tacit and explicit knowledge in a spiral of learning and knowledge development.

Dr Peter Smith, a principle research associate at Marcus Oldham College in Victoria, Australia, views the development of knowledge in the Thoroughbred Industry "*as the development from being a novice to being an expert through several stages*". He proposes that there are four types of knowledge that are vital to understand and develop in terms of expertly using knowledge in the Thoroughbred Industry (Morel and Smith, 2010).

Type of Knowledge	Characteristics of that Knowledge	Where that Knowledge is developed
Propositional	Knowing Theories and Facts	Mainly in education and training
Procedural	Knowing Processes and how to do things	Some education; mainly in the workplace
Strategic	How to decide what to do and when	Mainly in the workplace
Attitudinal	Workplace values and attitudes	Mainly in the workplace

Table 3 Types of Knowledge used in the Thoroughbred Industry (Dr Peter Smith)

Morel and Smith (2010) surmise that it is the procedural and the strategic knowledge learnt in the workplace that is really important in developing the skills and knowledge required to be a valued industry expert. The previously discussed Knowledge Industry Framework, combined with the invaluable knowledge gained by industry participants in the workplace, provides the guiding principles required in order to develop a Knowledge Management initiative, which can be used to organise and make available the codified knowledge within the industry.

2.3.2 Managing Knowledge

Traditional Knowledge Management (KM) is concerned with the development of systems used to collect and organise an organisation's knowledge in order to make the knowledge more useful to the organisation. The KM system should be continually evolving to incorporate new knowledge for the process of generating new knowledge and provide this knowledge in an organised manner that is easily accessible.

It should be noted that the majority of Knowledge Management literature is predominantly focused on the concept of Knowledge Management within organisations. While numerous comparable concepts exist, the concept of culture can vary quite differently between an organisation and an industry due to the removal of geographic restrictions. This is an important factor to consider when reviewing KM within a Knowledge Industry.

Collison and Parcell (2002) suggest that KM can be described as a combination of people, process and technology, with each being an important factor in a successful KM system; 70% People, 20% Process, 10% Technology. Knowledge Management can thus be defined as a collective approach by "*connecting the people who know via processes to simplify sharing with common reliable technology"*. More recently with the advent of Web 2.0 and Social Software, the concept of Knowledge Management has evolved into a community based discipline dependent on people participation and social inclusion.

In the context of this research, KM can be seen as managing information to utilise the knowledge in an industry or community in order to benefit from finding and applying innovative answers to old and new questions. KM does this by achieving the following (Awad & Ghaziri, 2003):

- Making community knowledge visible no matter where it is.
- Provides access to an industry's collective expertise anywhere in the industry.
- Retains the industry's knowledge in times of change.
- Exploits knowledge as a community asset.
- Helps to ensure the knowledge is up to date and relevant.

- Enables the survival of the industry.

Awad & Ghaziri (2003) further conclude that KM is shaped by social and human factors that require the involvement of the community. Communication takes place when individuals or groups are stimulated by personal and social motivators that encourage the sharing of experiences and observations. The true representation of industry knowledge is one where the community voluntarily explores, uses, and adopts knowledge in the best interests of the industry.

2.3.3 Industry-Wide Knowledge Sharing

"Human beings have an infinite ability to create knowledge. Add the convenient fact that unlike conventional assets, knowledge grows when it is shared, and you have the two most powerful features, which will change how we manage in the Knowledge Era." Karl-Erik Sveiby

Knowledge sharing is an activity through which knowledge (i.e. information, skills, or expertise) is exchanged among people, friends, or members of a family, a community or an organisation (Wikipedia A, 2011) Marsick and Watkins (1999) theorise that Knowledge Sharing can be seen as a type of informal learning where knowledge is imparted and obtained reflexively. This Informal learning relates to the lifelong process in which every person acquires and accumulates knowledge, skills, attitudes and insights from daily experiences and exposure to their environment.

This view of Knowledge Sharing is based solely on the concept of Internalisation of tacit knowledge, where people accumulate new tacit knowledge through the internalisation of conversation, observation and life experience. In the context of the Thoroughbred Industry this form of Knowledge Sharing can be seen in the master-servant relationships associated with apprenticeships and continuation of family legacy, as discussed in section 2.2.1.

Knowledge Sharing must also be defined in terms of Explicit Knowledge in order that a representative argument can be presented. Avram et al. (2005), posit that sharing explicit knowledge can be described as the acquisition and storage of knowledge in

knowledge bases. A number of the major Thoroughbred Sales Companies have embraced this concept of knowledge sharing by providing new technology which enables sales goers, view interactive sales catalogues and pedigree reports, make personal notes on electronic copies of sales catalogues, and share these notes with other sales goers.

In the context of a Knowledge Industry, the concept of informal learning is retained as the primary method of internalisation of tacit knowledge however sharing of explicit knowledge transcends the traditional use of central knowledge repositories. Instead a collaborative approach is taken where explicit knowledge is shared through the guise of a communal or social network, peered and maintained by the industry community.

2.3.4 Barriers to Knowledge Sharing

Within a Knowledge Industry the identification and removal of knowledge sharing barriers plays a pivotal role in the success of any Knowledge Management initiative. In the context of a Knowledge Industry, development of a knowledge sharing initiative must consider both the individual and social barriers, due to the global and often culturally diverse nature of the members of the industry. A number of the main individual barriers, within the Thoroughbred Industry, that can be identified are:

- "*Knowledge is power*" – this could refer to the owner/manager of a small stud farm, or a racehorse trainer, not wanting to lose the trade secrets that they attribute to their own unique way of achieving success.
- *Personal discovery* - People have pride in not having to seek advice from others and in wanting to discover new ways for themselves.
- *Lack of trust* – knowledge in the Thoroughbred Industry is viewed as a vital commodity when achieving success. Sharing this knowledge with others could provide industry competitors with an added advantage.
- *Lack of time* – due to the time and labour intensive nature of the Thoroughbred Industry, lack of time can be identified as a dominant barrier to knowledge sharing.
- *Conflict of motives* – this barrier can be best described as producing "win – win" or "lose – lose" situations. Industry experts are not willing to share the

knowledge attributing to their success for others to also achieve success. Alternatively, industry experts are also adverse to sharing bad knowledge as to do so, would also damage the expert's industry reputation.

- *Lack of Communication Skills* – very often, experts in any domain, although knowledgeable and skilled, can lack the communication skills required to articulate and express their knowledge in a way that can be shared with others.

In the context of a Knowledge Industry, the social barriers to sharing must also be examined to fully understand the potential for failure of an industry-wide knowledge sharing initiative. The following are some of the social barriers that can be considered.

- *Language* – due to the global scale of the Thoroughbred Industry and its traversal of all international borders, language can become an issue when sharing knowledge. Many sectors within the industry attempt to combat this by employing multilingual individuals who can converse and dictate globally.
- *Bureaucracy and Hierarchy* – high levels of bureaucracy and hierarchy within governing bodies can cause a bottleneck in the knowledge sharing initiative. In a number of areas these bureaucratic bottlenecks are being alleviated through the provision of electronic services.
- *Reluctance to embrace technology* - due to lack of familiarity and experience with certain technology, industry members may not be willing to impart their valuable knowledge via untrusted technologies.
- *Work Environment* – in the Thoroughbred Industry the physical work environment and layout of work areas restrict effective sharing of knowledge to other industry members not directly associated with that work environment.
- *Size of Industry* – as discussed earlier, the Thoroughbred Industry is a global operation, traversing both hemispheres and all international borders. The magnitude of a knowledge sharing infrastructure required to adequately service the global community could be considered as a barrier to the effective transfer of knowledge.

For many years the retention of tacit knowledge has been a key barrier to Knowledge Sharing in the Thoroughbred Industry. Breeders, Trainers, Agents, etc. have all in

retained personal tacit knowledge specific to their domains and roles making it difficult for that knowledge to propagate within the Industry.

Sharing Culture – Breaking Down Barriers

"*Sharing knowledge is not about giving people something, or getting something from them … Sharing knowledge occurs when people are genuinely interested in helping one another develop new capacities for action; it is about creating learning processes.*" Peter Senge

To promote the development of a knowledge sharing culture within the Thoroughbred Industry, it is important to identify the link between the culture, practices and behaviour of members in the industry and also the social barriers within the industry. Fahey and DeLong (2000) have produced studies showing that industry culture influences knowledge sharing by as much as 80%. Consequently, to effectively develop an industry-wide knowledge sharing culture, a change in the culture of the industry is required including; communication, trust and technology.

Due to the logistical difficulty involved in culture change within a global industry, there is no single identifiable area to start the process. Most culture change is based on a layered model approach which can be seen to address the underlying social barriers in order to facilitate change in the individual or personal barriers. This should be addressed as an industry-wide agenda in order to facilitate all areas and members of the industry, avoiding promotion of one group over another which could lead to further barriers, through breakdown of trust, conflicted motives and decreased belief in personal discovery.

This is particularly true in the context of the Thoroughbred Industry, as the knowledge contained within the industry does not lie with any one individual or group of individuals. Although large amounts of industry knowledge, resides with experts, the knowledge held by other industry participants cannot be easily discounted in their favour. For example, Racing Tipsters spend their career examining and analysing data relating to Horses and their racing patterns. The knowledge acquired from these exercises is then shared in order that 'punters' can make informed decisions, based on knowledgeable expert opinion, when backing Horses in races.

2.3.5 Communities of Practice (CoP)

Communities of Practice (CoP) are social structures that focus on knowledge and explicitly enable the management and sharing of this knowledge to be placed in the hands of the members of a community. Wenger (2006) indicates that Communities of Practice are communities "*formed by people who engage in a process of collective learning in a shared domain of human endeavour.*"

It can therefore be argued that CoPs are the principals of knowledge management within a Knowledge Industry due to their communal commitment to knowledge sharing in the pursuit of new knowledge. Since the main principle of a community of practice is group-related, like other knowledge sharing methodologies, it relies on the cooperation and collaboration of its participants. A successful CoP should represent a strong social, yet informal, forum which includes knowledge sharing, discussion on certain topics, problem solving, learning, and general interaction.

Wenger (2006) proposes the following concepts relating to a community of practice:

- A community of practice is not simply a group of friends, there has to be shared competencies among members as well as a shared domain of interest.
- Communities of practice require relationships, as relationships are the foundation of sharing knowledge and learning.
- Members of communities of practice are practitioners, i.e. consisting of a shared collective of resources, or a shared practice

Wenger's concepts give emphasis to the key attributes used to define a CoP including; the importance of a shared domain of competencies and interests, reciprocation of assistance encourages the willingness of participants to share knowledge and a commitment to contribute is required from all participants.

Three categories can be derived in the context of CoP in a Knowledge Industry:

1. *Knowledge Development* - A CoP provides participants with a medium for sharing "tricks of the trade" and best practices (Zárraga-Oberty & De Saá-

Pérez, 2006), by facilitating the flow of innovative knowledge sharing practices (Senge et al., 1999).

2. *Social development* - Allee (1999) indicates that companies, and therefore industries, "do not exist in a social or environmental vacuum". CoPs improve communication, consideration, and member relationships, which can be seen to improve and increase knowledge sharing within the community.
3. *Problem solving* – Ardichvili et al. (2003) maintain that communities of practice are based on problem solving with members of a *CoPs*, through communication, collaboration and alternative viewpoints, constantly finding new ways to manage and resolve problems relating to the industry.

When examined using these three categories CoP can be found to play an important role in the development of new knowledge through social development and interaction by means of extensive and collaborative problem solving. In the context of a Knowledge Industry, a revised definition of a Community of Practice might be given as, "*An informal community of like-minded individuals with a common interest and desire for sharing knowledge and the development of new knowledge, through social mediation and problem solving, with the intention of promoting, sustaining and enhancing their industry*"

2.3.6 Shift in Knowledge Sharing

Stephenson (2004) implies that learning and knowledge are distributed and ubiquitous by nature. By stating, *"I store my knowledge in my friends",* Stephenson equates the collection and sharing of knowledge to "*collecting people*". Further elaboration of this concept leads Stephenson to propose that "*a friend of a friend is a friend*", examining the potential for knowledge sharing by people through their trusted relationships. Therefore to be more effective, Knowledge Management solutions, including Knowledge Sharing, need to function in an increasingly decentralised and socially accessible approach, based on informal, loosely bound and distributed control.

According to MacManus and Porter (2005), information and knowledge is distributed across multiple domains in the form of microcontent. Microcontent, a term defined by Anil Dash in 2002 (Dash, 2002), refers to small information fragments that can stand

alone or be used in a variety of contexts, including Blog Posts, RSS Feeds and other social software technologies. MacManus and Porter (2005) further stipulate that no single centralised database or technology system is capable of capturing and distributing all the knowledge contained within an industry, but that new sets of technologies are emerging to aggregate and reproduce microcontent, specifically knowledge, in new and useful ways.

Downes (2005) has observed that the web has shifted from being a medium to transmit and consume information to a platform for knowledge creation, sharing and management. This next generation web first referred to as Web 2.0 by O'Reilly (2005), which encapsulates the concepts, technologies, practices and principles of Knowledge Management, has at its major component (Alexander, 2006), Social Software used for effective and efficient Knowledge Sharing within a Knowledge Industry. In the context of this dissertation, Social Software includes technologies such as Weblogs, RSS and most importantly social tagging. It should be noted, however, that Social software is not restricted to these technologies.

Dr Simon Livingstone, Principal at Marcus Oldham College specialising in Horse Business Management, states that "*New knowledge should develop new practice; and old practice should be challenged from time to time by new knowledge. Our graduates are entering an industry where the knowledge and skills have traditionally been forged through the work. While that develops sound skills and practices it can also mean we are slow to innovate – and innovation and response to change are features of twenty-first century business life that we just can't ignore. We have to get those features operating in our businesses as a part of the everyday enterprise culture*." (Morel and Smith, 2010)

This statement can be viewed as a confirmation that today's knowledge workers are being exposed to new innovations and changes in technology which provide mechanisms for Knowledge Sharing using Web 2.0, and more specifically Social Software. A general trend, or paradigm shift, can be seen in the context of Knowledge Management and Sharing, from the conventional centralised repository to a more conversational collaborative intelligence. This collaborative intelligence can be seen in

the emergence of a new type of CoP within Knowledge Industries, with the emphasis shifting from real or conventional CoPs to Virtual CoPs.

Virtual Communities of Practice (VCoP)

Brannigan (2010) observes that the increased spread of technology and globalisation has resulted in the emergence and development of Virtual Communities of Practice (VCoP). A VCoP, as opposed to a real CoP, implies that there are no physical meetings, between the members of the VCoP, and that Knowledge Sharing is performed through the technologies used by the VCoP.

In the context of a Knowledge Industry the key advantage of a VCoP over the traditional CoP, is that the VCoP is not bound by geographical and cultural limitations. Knowledge Sharing can be achieved through the use of Social Software associated with the Web 2.0, since communicating with others in a VCoP involves the creation of a "social present" (Tu, 2002).

While Web 2.0 technologies and the online environment facilitate the development of VCoPs, Kimble, Hildreth & Wright have found that facilitation of participation is central to the evolution and success of the VCoP (Kimble et al., 2001). The facilitation of participation is more difficult in a VCoP as real-time face-to-face interaction is not present as opposed to a Real CoP where personal interaction is vital. It can be observed from this that the provision of technologies that are quick and easy to learn are imperative for the success of a VCoP in a Knowledge Industry.

The Thoroughbred Bloodstock Industry can be regarded as a Community of Practice formed by numerous groups of business people, experts and individuals who share an interest and/or knowledge in the area of Thoroughbred Horses. More recently with the shift to a more social web and the leveraging of its Knowledge Sharing capabilities, the Thoroughbred Industry has begun to evolve into a more socially oriented Virtual Community of Practice. Although in its early stages this transition can be seen, for example, in the emergence of portals such as the Thoroughbred Bloggers Alliance (TBA), who describes itself as "*a diverse group of horse racing professionals and fans blogging about the sport we love*" (http://tbablogs.com/index.php?about). As well as the use of blogs the TBA use other social software mediums, such as Facebook, to

Share Knowledge and provide a forum for feedback from other members of the Thoroughbred Industry VCoP.

2.4 Conclusion

This chapter has presented the current Thoroughbred Industry as a Learning Industry, which uses its vast amounts of knowledge as a learning enabler, in the hopes of becoming a Knowledge Sharing Industry. The fundamental principles of traditional Knowledge Management are challenged and the link between the culture, practices, behaviour and the social barriers within the industry are identified, in order to be able to develop a successful Knowledge Sharing initiative.

Finally, the chapter presented a discussion on Communities of Practice which encapsulate the current paradigm shift, from the traditional centralised repository to a more conversational collaborative intelligence, indicating a new approach to Knowledge Sharing within the Knowledge Industry. Through the use of Web 2.0 and Social Software technologies, scope is provided for the concept of Virtual Communities of Practice, where members are not limited by the geographic and cultural boundaries of a traditional Community of Practice.

The next chapter will introduce the concept of Web 2.0 and Social Software and will examine how these technologies can be used as a medium for Knowledge Sharing within the Thoroughbred Industry.

3 WEB 2.0 AND KNOWLEDGE SHARING

3.1 Introduction

The previous chapter presented an overview of the Thoroughbred Industry as a Learning Industry and the processes involved in becoming a Knowledge Industry. Key to this transformation is Knowledge Sharing and how it is a cultural trait that is inherent in people, but in some cases can be nurtured. It also introduced two of the concepts associated with Knowledge Sharing within an industry: *Communities of Practice* (Physical Sharing) and *Virtual Communities of Practice* (Virtual Sharing) and the shift in Knowledge Sharing from a centralised repository to a conversational collaborative approach.

This chapter examines how the use of Web 2.0 technologies, namely Social Software, can be used as a medium for Knowledge Sharing. In the context of this research the characteristic knowledge contribution, sharing and collaboration features of Weblogs and their RSS technology as a Knowledge Management and Sharing Tool will be investigated.

3.2 Web 2.0

The W3C defines the World Wide Web as "the universe of network-accessible information, the embodiment of human knowledge" (W3C, 2011). The Web has evolved into a global electronic publishing medium providing its users with the ability to share and accumulate information and knowledge. The present first generation of the web has changed our daily practice, and these changes have become even more significant with the advent of the second generation of the web, or Web 2.0.

Web 2.0 is used to describe a new generation of Web services and applications with an increasing emphasis on human collaboration. The term Web 2.0 was defined by Tim O'Reilly, CEO of O'Reilly Media, Inc., in 2004. In 2007 Musser & O'Reilly further defined Web 2.0 as *"a set of social, economic, and technology trends that collectively form the basis for the next generation of the Internet—a more mature, distinct medium characterized by user participation, openness, and network effects"*. This definition

introduces the concept that diverse demographic, technological, and economic changes are driving Web 2.0. Underlying them is people's fundamental desire to connect, communicate, and participate (O'Reilly, 2006).

While this and the myriad of other existing definitions of Web 2.0 differ vastly, a common theme has emerged indicating that Web 2.0 is a people centric platform providing unparalleled openness for connection, collaboration and participation in Knowledge Sharing. Webopedia's definition of Web 2.0 further strengthens this stating that the "*improved functionality of Web 2.0 includes open communication with an emphasis on Web-based communities of users, and more open sharing of information*".

In 2005 Tim O'Reilly, following extensive discussion and brainstorming to derive the characteristics of Web 2.0, formulated the richer dynamic features of Web 2.0 over its more static Web 1.0 predecessor.

Web 1.0		Web 2.0
Britannica Online	→	Wikipedia
Personal websites	→	Blogging
Domain name speculation	→	Search engine optimisation
Publishing	→	Participation
Content Management Systems	→	Wikis
Directories (taxonomy)	→	Tagging ("folksonomy")
Stickiness	→	Syndication

Figure 7 Web 1.0 vs Web 2.0 (adapted from O'Reilly 2005b)

According to Berners-Lee, Hendler & Lassila (2001), the Web has become a core platform of communication, with special interest groups created in the form of Virtual Communities to share their common interests and knowledge. Virtual Communities try to exploit these richer features and forms of knowledge assets focusing on the social and collaborative dimension of Web 2.0 (Kirchner et al., 2009).

3.2.1 Characteristics of Web 2.0

Web 2.0 is both a platform on which innovative technologies have been developed and a space where users are treated as first class objects. The platform consists of various pre-existing and new technologies on which a variety of popular social networks such

as Facebook, etc. have been built. The principle characteristic of any of these social networks is that the participating users are as important as the knowledge content that they upload and share with other users of the community (Cormode and Krishnamurthy, 2008). Numerous technological aids have been created in Web 2.0, to increase the easiness and potential for knowledge creation, allowing anyone within the virtual community to contribute knowledge and generate new knowledge. – Any user can be a knowledge expert.

Web 2.0 provides a two-way interactive mechanism enabling Knowledge Sharers to contribute up-to-date knowledge to the shared community. The democratic nature of Web 2.0 facilitates the creation of niche knowledge communities who can exchange knowledge content, tag, comment and link to both intra-community and extra-community knowledge.

Lee and Lan (2007) have proposed the following list of characteristics as the objectives of knowledge sharing using Web 2.0.

- Contribution – every Internet user has the opportunity to freely provide their knowledge content to the relevant subject domains
- Sharing – knowledge contents are freely available to others. Secured mechanisms may be enforced to enable knowledge sharing amongst legitimate members within specific communities.
- Collaboration – knowledge contents are created and maintained collaboratively by knowledge providers. Internet users participating in the knowledge contents can have conversations as a kind of social interaction.
- Dynamic – knowledge contents are updated constantly to reflect the changing environment and situation.
- Reliance – knowledge contribution should be based on trust between knowledge providers and domain experts.

Web 2.0 can be seen as a shift from the individual to the community, replacing individual organisational knowledge with industry wide community knowledge, by providing **Reliable** and **Dynamic** knowledge **Content** that can be **Shared** and **Collaborated** on. These characteristics are promoted by the various technologies available with Web 2.0, collectively known as Social Software. These software

technologies will be further investigated, in Section 3.2.2, with a brief literature review of the positive and negative impact of Social Software. A more detailed examination of Weblogs will follow in Section 3.3.

3.2.2 Social Software

In recent times a paradigm shift in Knowledge Management, has been observed, from a centralised repository to a more conversational communal and collaborative approach. The growing phenomenon of Web 2.0 has augmented this with sets of tools and technologies that are simpler, transparent and more flexible, allowing users to impart their knowledge on their own terms. The technology and tools associated with Web 2.0 have recently been termed "Social Software" used to designate, "*the use of computing tools to support, extend, or derive added value from social activity - Including (but not limited to) weblogs, instant messaging, music and photo sharing, mailing lists and message boards, and online social networking tools*". (Avram, 2006)

Tom Coates (2003) posted an entry in his weblog, plasticbag.org, on May 8, 2003 in which he defined Social Software as *"a particular sub-class of software-prosthesis that concerns itself with the augmentation of human social and / or collaborative abilities through structured mediation (this mediation may be distributed or centralised, top-down or bottom-up/emergent)"*. This sparked a community debate and discussion, using a collaborative knowledge sharing approach, i.e. commenting, aimed at expanding and exploring various definitions of social software. In this way social software uses the web as a collaborative medium that allows users to communicate, work together, share and publish their ideas and thoughts as collective Knowledge.

With the advent of Web 2.0, Social Software provides users with technical transparency for conversations and collaboration, for knowledge creation, sharing and publication, for identifying experts and accessibility to expert opinions globally. Users of all technical ability now have complete control over the tools and methods for Knowledge Sharing within the global web community (Byrne and Goddard, 2010)

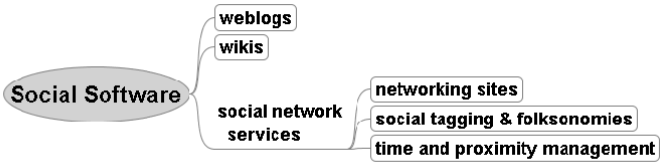


Figure 8 Key areas of Social software

Two of the more common types of Social Software are Wikis and Weblogs, as depicted in Figure 8 Key areas of Social software.

- Wiki - a website (or other hypertext document collection) that allows users to add content, as on an Internet forum, but also allows anyone to edit the content.
- Weblog - a web application enabling periodic posts on a common webpage with public access. (A more indepth examination of Weblogs will be discussed in Section 3.3)

The sudden popularity of social software technologies is attributed to the increase in low-cost tools and the critical mass of millions of people who are now connected to the Internet (Boyd, 2005), to the growing tendency of people to rely more on their own personal social networks than on traditional company structures (Schechter et al., 2009), and to the people's need to feel part of a community (Bryant, 2003).

3.2.3 The Good and The Bad

Within the existing literature there are two emergent schools of thought regarding the use of Social Software, the first exhibiting a fierce positivity towards the usefulness and benefits of Social Software, the second admonishing the openness of Social Software highlighting the negative implications of its use. Pressley (2006) conducted a literature review of existing writings to examine these two schools of thought identifying a number of existent trends within both.

The majority of the literature is in favour of using Social Software, extolling a number of positive trends associated with its use, including:

- *Allowing for Inexpensive Collaboration* - The abundance of free and open source Social Software provides a cost effective means of collaboration.
- *Allows for Efficient Real-Time Communication* - Social Software supports efficient, effective and real-time communication within a community.
- *Allows for Good Public Relations* - Community requested information and knowledge can be provided through blogs and wikis making use of Really Simple Syndication (RSS) to make this information conveniently accessible.
- *Allows for Online Archiving* - Christian Wagner (2005) investigated how some of the newer collaborative Social Software allows for free movement of knowledge providing the basis for online archiving of individual knowledge that might otherwise have remained tacit.

The potential drawbacks were also acknowledged within the literature, forming the negative evidence for the second school of thought. The negative trends include:

- *Potential for Misuse* - The strength of Social Software can also be its weakness, by providing increased levels of openness and transparency, leading to misuse.
- *Too Little Control* - Robert Niles (2005) indicates problems with the lack of control in Social Software, if community contributors can not be adequately controlled.
- *Lack Of Standards* - The oldest concern that has yet to be corrected is the lack of completed standards in Social Software. While much progress has been made in this area a number of non-standard aspects remain, including; blogging platforms, sharing resources and instant messaging formats to name a few.

Mejias indicates that the advent of Web 2.0 and Social Software signifies an important social turn in terms of new concern and interest in the social lives of other users. Focus must be given to the reality as opposed to the virtuality of the social interactions, in order to assess the benefits of Social Software (Mejias, 2005). Within social communities, there is a need for more attention to be paid to governance, shared values, rules and acceptable use to ensure that Social Software may be of benefit to all users. (Bryant, 2003)

3.3 Weblogs

Avram (2006) describes a weblog as a type of Social Software, enabling periodic posts of knowledge content on a common webpage with community access. These posts are usually, but not necessarily, in reverse chronological order, displaying the most recent posting first. Editing a weblog does not require any special training, enabling anyone in the web community to publish content on the web and in essence become a knowledge sharer. Originally weblogs were maintained by single authors, but the ideals of community based networking on the web has seen an increased shift towards portal based blogs where many authors share knowledge, e.g. http://www.kmblog.com/

The entries, called posts, are usually short with the most recent ones displayed on the weblog homepage, while old posts can be retrieved from archives ordered chronologically (and possibly by topic). Many posts link to interesting on-line articles, earlier discussions or related readings and enable readers and other weblog authors to add comments or link to a particular post using a static web address, known as a permalink (permanent URL).

As with any other website, the web community can use any HTML browser to visit weblogs making them a universal method of knowledge sharing across all platforms. Weblogs range from personal diaries to professional knowledge repositories maintained by industry experts. Many weblogs enable visitors to leave public comments providing a platform for community based collaboration and generation of new knowledge.

Most weblogs provide a facility for sharing its content through syndication of its blog posts using RSS (Really Simple Syndication, or Rich Site Summary). This enables readers to keep up with many weblogs (and an increasing number of other websites), without navigating the actual web pages (Efimova, 2004). RSS and Weblog sharing facilities will be discussed in more detail in Section 3.5 of this chapter. Weblogs also provide search tools allowing users find weblog posts or connections between them. Additionally, search engines like Technorati (http://technorati.com/) provide users with a means of searching for and finding blogs with a common field of expertise or knowledge.

According to Technorati, since 2004 the web has seen explosive growth and maturing of Blogging and Weblogs. Technorati's State of the Blogosphere series chronicles the rise and evolution of the Blogosphere as we know it. The blogosphere is made up of all blogs and their interconnections. The term implies that blogs exist together as a connected community (or as a collection of connected communities) or as a social network in which everyday authors can publish their opinions. The following statistics have been condensed from the "State of the Blogosphere 2009" published by Technorati.

- More than 133,000,000 blogs have been indexed by Technorati since 2002
- 77% of Internet users read blogs according to Universal McCann
- 60% are 18-44 * 75% have college degrees and 40% have graduate degrees
- 72% say they blog in order to share their expertise.
- 57% say that their future plans include blogging even more (including 74% of 18-24 year olds).
- Part-Timers, Pros, and Self-Employed Bloggers are blogging as much as or more than ever (73%, 76% and 80%, respectively), while Hobbyists are blogging somewhat less.
- 15% of Bloggers spend 10 or more hours each week blogging.
- One in five Bloggers report updating on a daily basis.
- The most common rate of updating is 2-3 times per week.
- The majority of blogs use tags (85%).
- 56% say that their blog has helped their company establish a positioning as a thought leader within the industry.
- 58% say that they are better-known in their industry because of their blog.

(Technorati, 2009)

Many weblogs also exhibit blogrolls, lists of weblogs that their authors read regularly which allow occasional readers find trusted "*sources*" that influence the thinking and writing of a particular weblog author. These links are not only referrals to specific sources, but also signs of value and personal recommendation. In this sense, hyperlinks between weblogs fulfil a similar function as references in scholarly publications (Mortensen and Walker, 2002).

3.3.1 Weblogs as a Knowledge Management Tool

Efimova (2004) hypothesises that knowledge management is unique and difficult to standardise and that the success of many Knowledge Management initiatives depend on the Knowledge Sharers willingness to adopt new practices and. Though the average public weblog is used by the author as a personal diary of their day to day thoughts, Efimova (2004) also observes that there are strong indicators that growing numbers of public Weblogs are used by professionals as Personal Knowledge Repositories and Knowledge Management Tools.

Weblogs can be used as tools for Knowledge Management by virtue of their nature as forums for presenting related knowledge and links, or references, to associated knowledge resources. Through these references readers can find trusted knowledge sources, which serve as a visible trace of the Knowledge Sharers expertise and acts as a starting point for initiating Knowledge Sharing connections within the community. Oravec (2004) noted that issues of blogger credibility may compromise the perceived validity of the shared knowledge, but this is outweighed by the inherent value of centralised access to community related resources.

By their very definition, Weblogs easily lend themselves to active Knowledge Sharing, through posting, collaboration, debate and argumentation. Mortensen and Walker (2002) suggest that Weblogs are instrumental in filtering and organising information and knowledge and sharing that same knowledge for the benefit of their users. Efimova (2004), opines that developing knowledge requires these filtering and organisation processes in order to make sense of knowledge contained within a Weblog. This supports the premise that Weblogs are useful as Knowledge Management Tools, but that both the Knowledge Sharer and the Knowledge Recipient are responsible for managing and filtering the Knowledge respectively.

3.4 Support for Knowledge Management

In their paper titled “Knowledge management(s)”, Despres and Chauvel (2000) proposed a framework for categorising Knowledge Management (KM) practices, taking into account five types of processes or activities:

1. Scan/map - pointing to the world of business intelligence, perception
2. Acquire/capture/create – associated with the world of research, development and creation
3. Package/codification/representation/storing – related to the world of databases, information and knowledge bases, organisational memory
4. Apply/share/transfer – related to the world of competencies, teamwork, intranets and cross border sharing
5. Reuse/innovate/evolve/transform – associated to the world of leverage, intellectual assets and innovation.

Avram (2006) postulated evidence that the use of Weblogs can support Knowledge Management and, more importantly, Knowledge Sharing can be seen when Weblogs are examined with respect to the aforementioned processes and contexts.

3.4.1 Scan/Map

Weblogs provide the facility for collecting information providing a map or web based yellow pages for locating and connecting with knowledge experts within the Weblogs domain. News and RSS aggregators provide a quick and easy way of subscribing to and monitoring the feeds provided by the weblogs of these knowledge experts.

3.4.2 Acquire/Capture/Create

Weblogs provide the capabilities for acquiring knowledge through integration of feeds, capturing knowledge by archiving and grouping and knowledge creation by individuals and groups by means of a posting-collaboration-debate-feedback mechanism.

3.4.3 Package/Codification/Representation/Storing

Packaging, codification, representation and storing of Knowledge are transparent processes closely associated with the acquisition, capture and creation processes of blogging providing a very intuitive and easy to use means of Knowledge Sharing within web based communities.

3.4.4 Apply/Share/Transfer

In Weblogs, Knowledge Sharing is implemented by providing RSS feeds containing information and links to the latest Knowledge and topics published on the Weblog. Users can acquire the shared knowledge by subscribing to any relevant RSS feeds within their domain.

3.4.5 Reuse/Innovate/Evolve/Transform

The main exponent of blogging is reflection, both on existing and newly acquired knowledge. This in turn can trigger a cycle of evolution and knowledge transformation as cross community conversations and debates are raised about individual or multiple Weblog posts. Avram (2006) points out that this phenomenon can occur across teams of Knowledge Sharers, across differing professions, countries and continents.

3.5 Sharing facilities (RSS)

RSS is an XML-based data format for websites to exchange files that contain publishing information and summaries of the articles posted to the Blog Site. In its first incarnation, RSS was defined as Rich Site Summary (Powers et al., 2002). As it has become more widely used for blog content syndication, RSS simply became known as Really Simple Syndication (Lee and Lan, 2007). A lot of blogging tools now create and publish RSS feeds automatically and webpages and blogs frequently display small RSS icons and links to allow a quick process of registering a feed from the site.

The RSS encoded content feeds can be harvested by news or RSS aggregators, which automatically check feeds for updated content and display new posts. This convenient method for monitoring large numbers of sources has led to a widespread use of RSS aggregators. (Klamma et al., 2007) The use of RSS is a very important accelerant in the knowledge retrieval process allowing users receive knowledge updates without interaction. Consequently, knowledge is retrieved when available and constantly updated with new knowledge which is relevant for the user. RSS technology can function seamlessly without ever being consumed, or internalised by users making it an ideal conduit for Sharing of explicit Knowledge.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
        <title>RSS Title</title>
        <description>This is an example of an RSS feed</description>
        <link>http://www.someexamplerssdomain.com/main.html</link>
        <lastBuildDate>Mon, 06 Sep 2010 00:01:00 +0000 </lastBuildDate>
        <pubDate>Mon, 06 Sep 2009 16:45:00 +0000 </pubDate>

        <item>
                <title>Example entry</title>
                <description>Here is some text containing an interesting description of the thing to be described.</description>
                <link>http://www.wikipedia.org/</link>
                <guid>unique string per item</guid>
                <pubDate>Mon, 06 Sep 2009 16:45:00 +0000 </pubDate>
        </item>

</channel>
</rss>
```

Figure 9 Sample XML based structure of an RSS feed

In some cases weblogs without RSS encoded feeds have a difficult time attracting regular readers. These weblogs do not conform to the ideal of Knowledge Sharing Weblogs put forward in this dissertation. In real terms weblogs that do not provide shareable content can be likened to individuals who are not forthcoming with sharing their tacit knowledge.

3.6 Conclusions

This chapter explored the many facets associated with Web 2.0 and its uses for Knowledge Sharing through the use of Social Software. Although software applications have been used for some time for Knowledge Management, the advent of Web 2.0 and its associated technologies has raised the bar, pushing the Knowledge Management principles and concepts from formal organisational methodologies to a more social community based interactive methodology. Weblogs strongly envelope the principle processes associated with Knowledge Management. They are strongly social websites providing easy and transparent means for the Acquisition, Packaging, Storing and Sharing of Knowledge to wide communities of web users. Although lack of standards and control provide the potential for misuse, most of the available literature chooses to focus on the positive implications of Social Software as a Knowledge Management Tool. Finally, a brief description of the RSS sharing/syndication facilities used by Social Software and Weblogs was presented. The next chapter will examine RSS and Portals, and how they are used in the domain Thoroughbred Industry. A review of the categorisation and classification techniques used by portals will be presented, which serves to highlight the progression from informal to formal semantics on the Web.

4 TOWARDS A SOCIAL SEMANTIC INDUSTRY

4.1 Introduction

Web 2.0 can best be described as the accumulation of Web-based collaboration technologies that let users easily publish and share knowledge. The widespread acceptance of these technologies has led to what is called the Social Web, a medium for online communication and collaboration. The use of RSS, syndicated from Web 2.0 technologies, as a Knowledge Sharing tool can be viewed as the driving component used by various types of Portals on the Web. Combined with the progressive nature of categorisation and classification techniques, from the informal (Social) to the formal (Semantic), the progression from an informal to a semi-formal Knowledge Sharing Industry could be aided by the development of Social Semantic Knowledge Sharing initiative through the use of Vertical Information Portals and a Domain-Specific Ontology.

This chapter will investigate the use of RSS as a Knowledge Sharing tool through population of the information content consumed by the users of web portals. The progression of user expressiveness, when classifying knowledge, from Informal → Social → Hierarchical → Formal, will be examined through the use of Tagging, Folksonomies, Taxonomies with an emphasis on Ontologies and the use of domain-specific ontologies in specialist domains and the difficulties associated with their development. The chapter will conclude with a brief critique of the use of Portals in the Thoroughbred Bloodstock Industry.

4.2 RSS Aggregation for Knowledge Sharing

As discussed in the previous chapter, the use of Really Simple Syndication (RSS) is a very important accelerant in the knowledge retrieval process allowing users receive knowledge updates without interaction. Consequently, knowledge is retrieved when available and constantly updated with new knowledge which is relevant for the users.

The impersonal nature of RSS and the ability of any user to subscribe to a feed is of particular relevance to Knowledge Sharing in an industry, since experts can share their

knowledge with a wider community without the requirement of knowing the individuals in advance. In her article "RSS: The Next Killer App For Education", Mary Harrsch (Harrsch, 2006) proposed a number of ways that RSS feeds can be useful in the context of Knowledge Sharing.

- Subscribers can keep track of new Knowledge posted on favourite websites from a single convenient location, without any interface clutter.
- Using RSS and some form of aggregation tool the subscribers can be notified of updates from many websites that are of interest. Instead of re-visiting all of these sites on a daily basis, using RSS technology subscribers can easily track new Knowledge on each of the sites of interest.
- RSS provides a more inclusive and protected solution as it does not require users to reveal their personal details. It also means that anyone can subscribe to the RSS feed without the requirement of being included within a specific group or clique.
- Using RSS feeds, subscribers can stay informed of calendar events planned by Industry Organisations. This helps the subscriber avoid potential conflicts of interest, e.g. Racing Calendars, Sales Dates, etc.
- Domain experts can establish mutual RSS feeds relating to specific areas of interest. This allows access to up-to-date content on a wide variety of subjects automatically.

Harrsch further discusses how technologies such as web portals often offer targeted collections of links for domain experts and users. Portals allow users to create dynamically updated, personalised and socially connected web pages. Platforms such as PageFlakes, NetVibes, or iGoogle allow users/authors to drag and drop RSS feeds and other widgets onto a web page to create a starting, or home, page. PageFlakes and NetVibes also allow users to share pages with others. Through the use of RSS users can have information 'pushed' to them that is customised to their personal interests. RSS use in this context can be viewed as the driving component used by various types of Portals on the Web.

4.3 Portals

Portal is a term, generally synonymous with a gateway, for a World Wide Web site that is or proposes to be a major starting site for users. Using portals knowledge is obtained by the users when they get connected to the Web or a community. O'Leary (2000) proposed that Portals have become one of the biggest web trends, with organisations setting up portals to attract customers to their websites and, in some cases, to use them as mediums for Business-to-Business (B2B) ecommerce tools (O'Leary 2000).

The governing principle behind successful portals is that the web-based knowledge and services being presented as part of the portal should reflect the resources and services required by, chosen by and expected by the end user. The portal should monitor the usage of the end user and have the capability to suggest new resources and services that match the current interests of the user. The ability to discover, suggest and aggregate various knowledge services and resources and make them available to the end user makes portals an important tool in Knowledge Sharing on the Web.

Most definitions of portals depend on the context in which they are being applied, however in the context of this research an investigation of the available literature on Web Portals revealed that a 'Portal' is:

- An application that provides a personalised and adaptive interface enabling people to discover, track, and interact with other people, applications, and information relevant to their interests (Liu et al., 2003).
- A website that provides an entry point to the Internet, and offers value-added services such as directories, searching, information news, and links to related websites (Rowley, 2000).
- A "Supersite" on the Internet that provides a comprehensive entry point for a huge array of resources and services (Han, 2003).
- A web address destination to which people with some common set of characteristics, interests, or needs go to gather information, interact with data, experience entertainment, exchange thoughts, or conduct transactions (Hurwitz, 2001).

Although the literature, does not agree on a single definition for a Portal, a common theme is that a portal is perceived as a means of discovering relevant knowledge of interest to the user. From the above definitions it can be concluded that a portal is seen either as a software application, or an Internet website used as an entry point to knowledge resources.

According to Kreuder (2002) the main driving forces behind a successful portal infrastructure are:

- *Information management* - The methods and procedures that are formulated and codified within the information management approach define the method of the development of the infrastructure.
- *Knowledge management* - Regarding the rising importance of knowledge, the theories and principles of knowledge management have strong influence on the portals design

Further literature reviews on portals indicates that there are several types of portals but no accepted system for classifying these portals. However, a number of broad categories of portals emerge from the literature that is of relevance to this research: Horizontal or Web Portals (Hortals), Vertical Portals (Vortals) and Vertical Information Portals (VIPs). It should be noted that sometimes a suitable portal implementation may require an integration of multiple types of portals implemented as a hybrid solution.

4.3.1 Horizontal (Web) Portals

Wikipedia describes web portals as a links page that presents information from diverse sources in a unified way. These portals are designed to act as gateways to the Internet, and they provide access to a wide range of web-based resources and services including: standard search engine features, e-mail, hosting, news and entertainment. Web portals are also referred to as horizontal portals as they provide general information and services covering many areas and domains which are of interest to a general audience.

Bertland described how a large number of public web portals are sponsored by the major search engine and browser providers, and may include alliances with other major players on the Internet (Bertland, 2001). One of the main disadvantages of these sponsored web portals is that they often refer users to resources, services and information that is owned or maintained by the same company, rather than to relevant impartial resources. Such examples include:

- Googles iGoogle web portal
- Microsofts MSN internet portal
- Yahoo's integrated internet portal

In the context of access to specialised domain knowledge, web portals end up overwhelming the users by trying to provide all types of knowledge. As a result another type of portal, the vertical portal, has emerged as a viable and compelling response to meet the requirements of increasingly specialist internet users.

4.3.2 Vertical (Domain Specific) Portals

While Web portals can be viewed as their immediate ancestors, vertical portals are designed to have a more focused knowledge domain area aimed at a specific particular category of users. Peek (1999) indicates that identifying portals that qualify as a vertical portal can be difficult but identifies a number of features that makes a portal a vertical portal:

- They should offer more than just content and have something that differentiates them from being virtual versions of magazines.
- Unlike web portals, they do not require a search engine to the outside world.
- They belong to the broad web community.
- They have clearly defined focus in terms of content and target audience

Vertical Portals concentrate on providing specific information content to their defined users and therefore do not include more general features such as email services, weather reports, and many others. Since vertical portals focus on defined user, in the

past they rarely provided facilities for personalisation. Many newer vertical portals have emerged that allow users a limited amount of personalisation, enough to tailor the content or theme for their viewing experience. This lack of complete personalisation and control is still a major disadvantage of the vertical portal model for knowledge management.

4.3.3 Vertical Information Portals

A vertical information portal (VIP) is a specialised entry point to a specific domain and or industry niche. Vertical Information Portals provide news, editorial content, digital publications, and e-commerce capabilities but extend the traditional vertical portal model, by also providing dynamic multimedia applications including social networking, video posting, and blogging. Information Portal refers to a type of portal which is implemented, not for e-commerce, but instead to provide access to digital knowledge resources and services for the community.

One of the most important beneficial features of Vertical Information Portals is the potential for increased collaboration within the targeted user community. In effect, increased collaboration translates to greater social integration of users across geographic barriers, providing the impetus for increased knowledge sharing.

4.3.4 Features of Portals

Much research has been carried out in order to define a common set of features that a good portal should possess. Ethridge, Hadden and Smith (2000) have identified and discussed several distinguishing features of portals as follows:

- Portals facilitate the discovery of people, organisations, and content in a meaningful context.
- Portals are secure, offering user authentication, credential mapping, and sensitive data encryption.
- Portals are personalised, proactively providing customised desktops based on the user's role in the community or enterprise.
- Portals are adaptive.

- Portals are single points of service, that provide a framework for accessing multiple heterogeneous data stores including enterprise databases, e-mail, and other multimedia resources.

In a recent paper, "Building Successful Portals", Joe Lamantia (2009) identifies six unique distinguishing features which, implicitly or explicitly, incorporate the features and characteristics identified by Ethridge, Hadden and Smith (2000). Ironically, the first letters of each of these characteristics spell the word PORTAL.

- ***P****ersonalisation for end users* – Personalisation is one of the most critical features of any type of portal as it establishes a relationship with the user, by allowing the user personalise the portal appearance and content to suit their individual requirements.
- ***O****rganisation of information*- Portals help to organise the user's information enabling quick and easy access to content when it is required.
- ***R****esource division* - Portals should separate facilities and privileges into various layers for different categories of users. For example, moderators should be the only people accessing features relating to the administration of the portal.
- ***T****racking activities* - Portals should be able to track user activity, interests, and behaviour in order to present users with a personalised view of information resources.
- ***A****ccess and display of information* - Portals should be able to provide users with access to information from multiple heterogeneous sources, including: Web servers, News feeds, Blogs and Wikis
- ***L****ocation of important people and things* - Portals must ensure that its users can locate the people and information they require quickly and easily. There must be ways to locate experts, communities and content related to a particular topic.

In light of the discussed portal features an appropriate definition of a Portal might be: "*A **P**ersonal, web-based application for **O**rganising information and **R**esources, enabling users to **T**rack, **A**ccess and **L**ocate experts, users and communities with common interests in a specific domain.*" This new definition attempts to encapsulate all the important features of a portal in the context of a knowledge industry as examined in this dissertation.

4.4 Tagging

An important feature provided by Social Software, Weblogs and Portals is the provision of a tagging system. This allows users to tag Web Pages, blog entries, photos, videos and any other available media so that they can be organised and later retrieved as well as shared with others having a common interest in that knowledge. Salari (2009) defines tagging as the process of annotating a web resource with a set of keywords. Terms used as keywords must be descriptive in order to organise content to aid with navigation, filtering, searching and retrieval.

Al-Khalifa and Davis (2007) have categorised tags into three categories:

- Personal - tags that have an intended audience and are often used to organise a user's own resources
- Subjective - tags that express people's opinions on the bookmarked web resource
- Factual - tags which identify 'facts' about the described web resource such as people, places, or concepts

Tagging, as discussed above, is the act of organising and making sense of many discrete, varied items according to their meaning. Further research reveals tagging to be a powerful tool for social navigation as it aids the sharing and discovery of new knowledge contributed by other domain members. Weller (2007) claims that a social community can benefit from tagging as it provides alternative views of the knowledge within the community. These views are reflected without cultural, social, political or personal bias allowing niche interests to be catered for such as the Thoroughbred

Bloodstock Industry where very specific tags are used that would not otherwise be used or required in other domains.

Halpin et al. (2007) indicate that there are both benefits and drawbacks to the tagging approach. The main benefit of tagging is that it allows much greater flexibility, see Figure 10, and adaptability in organising knowledge. Users are not constrained by a centrally controlled vocabulary which would restrict their choices of tags to existing words or phrases selected by authorities or a governing expert.

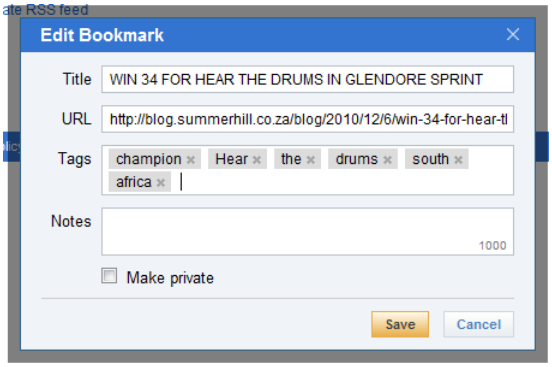


Figure 10 Example of flexible tagging system used by Delicious

The openness of tagging systems also contributes largely to the issues inherent in their use. The lack of a centrally controlled dictionary gives rise to numerous linguistic issues including:

- Polysemy – words that have multiple related meanings, e.g. "Class"
  - A room used for educational purposes.
  - The classification of a Thoroughbred Horse.
- Synonymy – different words that have the same meaning, e.g. Colt, Gelding, Filly and Mare are all synonyms for the word "Horse".
- Homonymy – words that have multiple unrelated meanings, e.g. "Horse" and "Hoarse".

An additional issue directly related to lose control of collaborative tagging is known as the "Vocabulary Problem" or variability in word usage among users. Many users may use different words or tags to describe the same things, which can lead to missing or irretrievable knowledge via conventional searching or browsing. This loss of knowledge can be counteracted through use of an intelligent tag searching facility, as

used by the Community based bookmarking portal, e.g. http://www.delicious.com. An additional attempt to counteract the search issues related to tagging has been to provide users with a suggestive starting point through the use of tag clouds.

4.4.1 Tag Clouds

Defined as "*a set of words, typically a set of tags, in which attributes of the text such as size, weight or colour can be used to represent features (e.g., frequency) of the associated terms*" (Halvey and Keane, 2007), the basic idea of a tag cloud is to represent tags according to their meaning, their weight and their frequency relative to other tags. This provides users with a visual facility that can be used to help identify and retrieve knowledge based on the most common and most relevant tag. Often, more frequently used tags are depicted in a larger font or otherwise emphasised. Selecting a single tag within a tag cloud will lead to a collection of items that are associated with that tag e.g., Blog Post, Photo, Article, etc.

The first widespread use of tag clouds was on the photo sharing website Flickr, but in more recent times they have also been popularised by Del.icio.us and Technorati, among others. The use of Tag Clouds can be used to discover trends in what people are tagging, for example, interesting links on Del.icio.us or cool pictures on Flickr. These are trends of popularity and tag clouds can be used as a way of depicting them.

There are three main applications of tag clouds in social software, distinguished by their functionality rather than appearance. The Tag size is indicative of a different method for analysing the tag usage in each case.

- The size of the Tag represents the number of times that tag has been applied to a single item.
- The size of the Tag represents the number of items to which a tag has been applied, as a presentation of each tag's popularity. This is the most common type, being used by portals such as Flickr and the blog aggregator Technorati.
- The size of the Tag represents the quantity of content items in a category represented by the tag.

In 2009, Lohmann, Ziegler and Tetzlaff conducted research on the usability of Tag Clouds and their layouts, based on the "Task-Related Performance and Visual Exploration" of users. The following are the research findings on tag cloud perception and performance: (Lohmann et al., 2009)

- Tag size: Large tags attract more user attention than small.
- Scanning: Users scan rather than read tag clouds.
- Centring: Tags in the middle of the cloud attract more user attention than tags near the borders (depending on the Tag Cloud layout).
- Position: The upper left quadrant receives more user attention than the others (based on Western reading habits, left to right).
- Exploration: Tag clouds provide suboptimal support when searching for specific tags (if these have not a very large font size).

Joe Lamantia (2009) indicates that during 2006, tag clouds moved beyond their well-known role as navigation mechanisms and indicators of activity within social media experiences, emerging as a standard visualisation technique for texts and textual data in general.

Figure 11 Sample Text Cloud generated from Summerhill Blog RSS feed

Lamantia (2009) refers to these new visualisations as text or word clouds, see Figure 11, indicating that they are visualisations of word frequency in a given text as a weighted list as opposed to tag frequency associated with the text. The emergence of this new form of text cloud is of particular importance for Blogs or Portals where users may not actively Tag their posts or feed content.

4.4.2 Folksonomy

The sets of categories derived from tagging are commonly referred to as folksonomies, describing "*the users, resources, and tags, and the user-based assignment of tags to resources*" (Hotho et al., 2006). Since a folksonomy is not bound by a hierarchy or any directly specified parent-child or sibling relationship between the tags they offer flexibility and adaptability in organising knowledge. However, some community consensus regarding a tags meaning is required so that users can label similar content with the similar tags.

Mikroyannidis (2009) suggests that folksonomies suffer from limitations such as ambiguity in the meaning of tags and lack of semantics, i.e. synonyms (student or pupil). Since the tag contributors of the folksonomy do not adhere to or operate using a centralised or controlled vocabulary, they can produce categorisations that are inconsistent or unrelated resulting in disparate and disconnected categorisation of knowledge. Recent studies have shown, however, that despite their distributed nature and lack of central control, folksonomies formed by collaborative tagging tend to reach a stable form under certain conditions depending on the number of users and the duration of the tagging process (Halpin et al., 2007). These findings are particularly important as a stable distribution is essential for accurate categorisation of knowledge.

Although not inherently available within published Weblogs, the ability for blog readers to tag blog entries is generally available within RSS feed readers or aggregators. Most RSS readers, e.g. Google Reader, BlogBridge, Feedly, etc. provide the blog reader with the ability to tag or retag certain blog posts and share them via their own blogs, tweets, email, etc. This functionality is a core attribute associated with the use of Portals and consequently provides for the concept of folksonomies within a community.

4.5 Taxonomy

The term Taxonomy is used to define a classification scheme with a hierarchical structure. Eleanor Rosch defines taxonomy as "*a system by which categories are related to another by means of class inclusion.*" (Rosch et al., 1976). A typical example of taxonomy, according to Gruber (2007) is the Dewey Decimal System (a

representation of books available in Western libraries and a means to efficiently structure those books), which is used by librarians to classify books according to a fixed categorisation scheme and organise them into shelves.

Taxonomies enable the representation of a set of concepts which can be used to model a domain of knowledge (Gruber, 2007). They establish a consistent means for the capture, use and re-use of knowledge within the specified domain. Taxonomies also serve the purpose of bridging the gap between formal semantics and informal tagging, allowing people operating in the same domain accept and comply with a common taxonomy and use the same definitions.

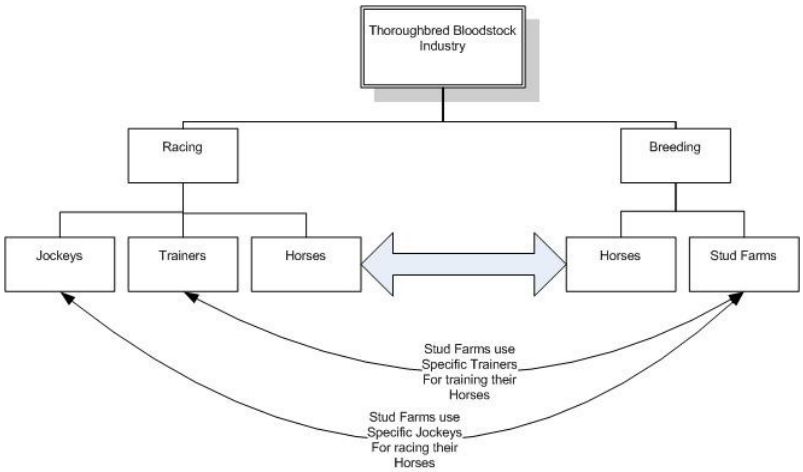


Figure 12 Example of hierarchical nature of Thoroughbred Bloodstock Industry

Figure 12 Example of hierarchical nature of Thoroughbred Bloodstock Industry, displays a sample of the hierarchical nature of the Thoroughbred Bloodstock Industry representing the progression of classifications and categories. It also highlights common classifications between the different areas. In the Thoroughbred Industry similar classifications can be used to categorise Horses in both Racing and Breeding operations.

Horses					
Stallion	Mare	Colt	Filly	Gelding	Foal

Table 4 Sample Classification of Horse

Taxonomy in the Thoroughbred Bloodstock Industry could also be used to indicate that racehorses are a type of Horse specific to the Racing area of the industry, e.g. 2 Year Old, 3 Year Old, etc. Additional classifications may be used to help determine the

types of races in which a racehorse can partake based on their racing experience, e.g. Maiden, Stakes Winner, Group Winner, etc.

In order to be effective the search and categorisation technologies within an information portal should be linked to a well-defined domain taxonomy. Using a well-defined hierarchy of subjects (taxonomy) facilitates easy retrieval and identification of knowledge within the portal.

4.6 Ontology

Ontology refers to the "*theory of existence*" in Philosophy, but has been adopted by the computer and information community to define a "*formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts*" (Gruber, 1992), which may be formally represented in a computer readable and usable format (Wikipedia C, 2011). The term has been widely employed since 2001, when Berners-Lee et al. envisaged the Semantic Web, which aims to turn the information stored on the Web into formally structured knowledge and according to Gruber (1992), produce an explicit "*specification of a conceptualisation*".

Park, Oh and Ahn (2010) discuss how ontologies provide a knowledge-sharing infrastructure supporting the representation and sharing of domain knowledge. They also indicate that an increasing number of ontologies are being developed, and their reuse and sharing offer several benefits by significantly saving the time and effort required building new ones.

An upper ontology (or foundation ontology) is a model of the common objects that are generally applicable across a wide range of domain ontologies. It employs a core glossary that contains, the terms, and associated object descriptions, as they are used in various, relevant domain sets. Ontology is a crucial component in knowledge management yet despite the widespread research and applications of general heterogeneous ontologies, developing domain ontologies remains a challenging issue.

4.6.1 Domain Ontology

A domain ontology (or domain-specific ontology) attempts to model a specific domain, or subset of a more general domain. It represents the definitions of terms and concepts as they apply to that specific domain. For example the word “Trainer” has many different meanings. An ontology describing the domain of Thoroughbred Horses would model the “Horse Trainer” meaning of the word, while an ontology describing the domain of clothing would model the “Running Trainer” (athletic shoe) meaning.

Since domain ontologies represent very specific concepts related to a very specific domain, they are often incompatible with other domain ontologies. Chen, Bau and Yeh (2011) indicate that this can present significant challenges to the ontology designer and developer, when deriving new ontologies from extant ones, as key terms may overlap and conflict with each other, while other specific terms may not fit in and will only serve to oversize the newly created ontology. Domain ontologies that use the same foundation ontology to provide a set of basic elements with which to specify the meanings of the domain ontology elements can be merged automatically. Ontology development for underdeveloped or specialist domains, which have no foundation ontology in existence, becomes even more difficult and time-consuming.

4.6.2 Development of Domain Specific Ontology

According to Pouchard et al. (2000), ontology development, or ontology engineering, is a discipline associated with knowledge engineering that studies the methods and methodologies for developing and building ontologies. The aim of Ontology development is to make explicit the knowledge within software applications, and within the industry processes for a particular domain.

Although some generic ontologies exist and are in widespread use, such as WordNet (a large lexical database of English - http://wordnet.princeton.edu/wordnet/) and top-level Cyc ontology (a comprehensive ontology and knowledge base of everyday common sense knowledge - http://www.cyc.com/), most applications require domain-specific ontologies that capture the terms and relations in a particular domain.

Noy and McGuinness (2001) have proposed a seven step development methodology to aid the process of a first time ontology development.

1. Determine the domain and scope of the ontology
2. Consider reusing existing ontologies
3. Enumerate important terms in the ontology
4. Define the classes and the class hierarchy
5. Define the properties of classes (slots)
6. Define the facets of the slots
7. Create instances.

Zhou, Booker and Zhang (2002) have identified a number of major issues associated with domain specific ontology development including:

- Construction principles – principles of development that are not clearly defined.
- Modelling tools – modelling tools with limited scope for domain specificity.
- Knowledge acquisition – difficulty in acquiring specialist knowledge from domain experts.
- Human Effort - ontology development consumes extensive human efforts.

Although research has shown that numerous small domain specific ontologies are available, most are focused on either established domains such as medicine and science (including computer science), or structured domains such as manufacturing. Many other underdeveloped domains have not been investigated yet and the identified issues become even more severe in such domains, i.e., the Thoroughbred Bloodstock Industry. They characterise an underdeveloped domain as a domain where:

1. There are no structured or semi-structured domain resources in existence.
2. There is no well-recognised authoritative information source to obtain all the information.
3. The contents are very broad.

It can be seen that one of the key factors in development of domain-specific ontologies is that it requires domain experts' involvement. This involvement can be in an individual or collaborative capacity aided by the assistance of a customisable collaborative system. This concept will be further investigated in Chapter 6 of this dissertation.

4.7 Critique of Portal use in the Thoroughbred Bloodstock Industry

The development of vertical portals may be compared to the development of specialised publications in the publishing industry. The industry started out with a handful of publications and now has a multitude of publications that target specific groups, including the Thoroughbred Bloodstock Industry. For example, people interested in Thoroughbred Horses read publications like The Racingpost, Pacemaker, Thoroughbred Breeder, Sporting Post, etc. to keep informed of the latest information in the industry.

Thoroughbred businesses, including Stud Farms, Feed Specialists and even IT companies, place advertisements in these specialist publications as they know that their target clients read them. Many of these media publications have also adopted the use of vertical portals to make their publications more accessible to the wider worldwide community. The Racing Post Portal, Figure 13, can be viewed as one of the leading Thoroughbred Industry based vertical portals, predominantly driven and populated with Thoroughbred News and Information it is also a source of diversified information for other sporting disciplines, emulating the success of its printed counterpart.

Figure 13 The Racing Post Portal

The Thoroughbred Bloodstock Industry has begun to embrace the great potential for electronic commerce that vertical portals offer. Many industry publications have adopted to use e-formats (Electronic Formats) instead of traditional media as a means of accessing a wider audience. Other Vertical Portals exists that have been designed to target specific subsets of the Thoroughbred Bloodstock Industry, see Figure 14. Such portals provide information and knowledge specific to the targeted domain participants including Racehorse Trainers, Breeders, etc.

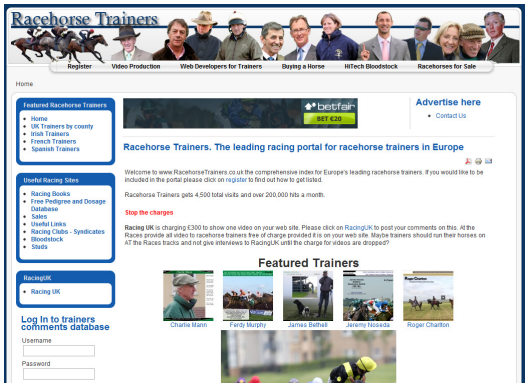


Figure 14 Racehorse Trainers portal - specifically for European Horse Trainers

One of the major industry problems is fragmentation of knowledge resources across the global domain of the Thoroughbred Bloodstock Industry. Collaboration through information portal structures could provide a means to establish common information and knowledge resources providing a universal and associated knowledge sharing effort in the Thoroughbred Bloodstock Industry.

Although collaboration, in this context, can be viewed as a positive move towards integration in the community, it is important to note the potential negative effect that such collaboration can have on the formal structures within the domain. For example, cross-hemisphere and even cross-continent structures can vary when referring to industry related concepts. Bloodstock Breeding Seasons differ between the Northern and Southern Hemispheres which can lead to potential confusion unless properly specified.

Another example could be the difference in Horse Ageing schemes between Hemispheres and in some cases within individual countries, e.g. Colombia, where Horses are aged differently depending upon the part of the country they were born in.

In truth the increased opportunity for collaboration is not always advantageous, and careful analysis is needed in order to validate the prospective benefits of the Vertical Information Portal model.

Despite these difficulties, vertical Information portals could be very useful in the Thoroughbred Bloodstock Industry, especially in areas that have a requirement for exposing and sharing knowledge for the benefit of all users.

4.8 Conclusion

Given the progressively social nature of the Web and its potential for Knowledge Sharing, the concept of Portal use and their use of RSS can be seen as the driving component for sharing knowledge across large communities. Various types of Portals are implemented based on the specificity of the knowledge domain and the requirements of its users. In the context of this dissertation, portals are considered to be potential tools for digital knowledge management and for providing access to well-organised, relevant information resources and services to end-users. Organising knowledge can range from the informal to the formal, or in the context of the Web, from the Social to the Semantic. Successful organisation of knowledge is paramount to adequately exposing it and ensuring that it is shared within the community through the use of structured or formal representation and the use of domain-specific ontologies. Although it is perceived that the creation of ontologies is a top-down and complex process, in reality ontologies can emerge bottom-up and be simple, based on the social tagging processes within a domain. This concept will be discussed and examined in Chapter 7.

The use of Portals for knowledge sharing in the Thoroughbred Bloodstock Industry was critiqued serving to highlight the types of Portals used and how information portals could be very useful in the industry for exposing and sharing information and knowledge for the benefit of all users. The next chapter will present the results of an industry survey aimed at evaluating the use of Weblogs, as a Knowledge Sharing tool, among Breeders, Trainers and Bloodstock Agents.

5 SURVEY ANALYSIS

5.1 Introduction

As previously mentioned in Section 2.2, for many years the retention of tacit knowledge has been a key barrier to Knowledge Sharing in the Thoroughbred Industry. Breeders, Trainers, Agents, etc. have all retained personal tacit knowledge specific to their domains and roles making it difficult for that knowledge to propagate within the Industry.

However, with the advent of the web and more importantly the widespread adoption of the usage of blogs within the Industry, more and more Industry Knowledge is being imparted and externalised. Faster and more efficient web access also provides industry members with a far greater medium for knowledge sharing by virtue of increased communication and collaboration. This form of knowledge sharing displaces previously imposed limitations that plagued the industry, such as global and cultural diversity, ageism and sexism.

This chapter describes the research undertaken to gain a better understanding of the attitude of the Thoroughbred Industry in respect of blog usage, for both internalisation and externalisation of knowledge. It has also been conducted in order to gain a better insight into the use of blog technologies in the Thoroughbred Industry, as a key enabler for the purposes of knowledge sharing.

5.2 The Survey

In order to better understand the impact and use of weblogs and online news sources within the Thoroughbred Industry, a survey was conducted through an online questionnaire. The survey was sent electronically, via email, using the www.kwiksurveys.com and the responses were collated through monitored feedback on the survey website. The reason for selecting an online survey was to provide a larger geographical distribution than could be achieved through standard postal surveys and face-to-face interviews. The greater geographical distribution would have

the advantage of being a truer cross-sectional representation of all members of the Thoroughbred Industry, as opposed to a subgroup limited by locale.

5.2.1 Structure of the Survey

The survey is divided into three main sections; the first to assess the demographics of the survey's respondents, the second deals with Blog use and the attitudes towards blogs, and the final section deals with the blogging activities of the respondents.

Two types of questions have been used throughout the survey, the first type of questions are direct questions, which the survey participant can respond to with simple Yes/No, or factual answers. The second type of questions use a Likert scale (Likert, 1932) to determine the strength of participant agreement or disagreement in response to a series of statements concerning their attitude toward blog use and benefits

5.2.2 Survey Participants

The survey was distributed to 300 Thoroughbred Industry professionals, which included a representative cross-section of 150 Stud Farms, 100 Trainers and 50 Bloodstock Agents from both the Northern and Southern hemispheres. The survey was complete with responses from 48 participants, from 280 surveys that were delivered (excluding 20 mail delivery failures), representing a total response rate of 17%.

By virtue of 11 years of working experience servicing the Thoroughbred Industry and significant knowledge of the mentality of its members, a number of factors have been identified which have contributed to the low response rate including:

- Lack of interest in survey participation
- Lack of technical knowledge
- Limited technical resources
- Limited time

Limited time was cited in a number of cases by respondents as a leading reason for not participating in the survey. This serves to highlight the labour and time intensive

nature of the industry, where Breeding and Training schedules contribute to a busy lifestyle with, in some cases, limited leisure time.

Although low, the response rate still represents a sufficiently large figure to representative of the industry and is indicative of a perceived lack of technical knowledge, trust and understanding in the context of blogs and their use.

5.3 Analysis and Display of Survey Results

From the onset of receiving completed questionnaires, the results were displayed and presented within the survey results section of the website (www.kwiksurveys.com). However in order to produce plotted charts and graphs, the final results for each question was entered into a master spreadsheet document in MS Excel. Each question in the survey was allocated a separate worksheet in the document.

During analysis the results from direct questions have been displayed as pie charts, whilst the responses to the Likert scaled questions have been analysed by subject in bar charts. The order of the analysis of the questions has been changed to that of the survey; this was done purposely, in order to first assess the range of respondents, secondly to review the impact of blogs on the participants and finally to assess the perception of the use of blogs as mechanism for knowledge sharing within the industry. For the purpose of this dissertation, the analysis is grouped into the following sections:

- Industry Demographics
- Blog Users Vs Bloggers
- Purpose of Blog Use
- Perception of Knowledge Sharing

5.4 Industry Demographics

This group of questions are focused on establishing the demographically diverse distribution of survey respondents. As previously discussed, a random distribution of industry participants was selected and invited to participate in the survey. The following demographical results provide an indication of the participatory distribution.

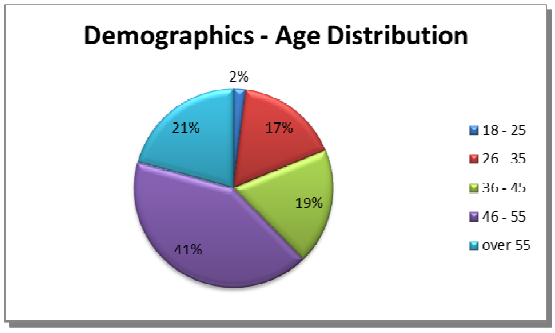


Figure 15 What is your age?

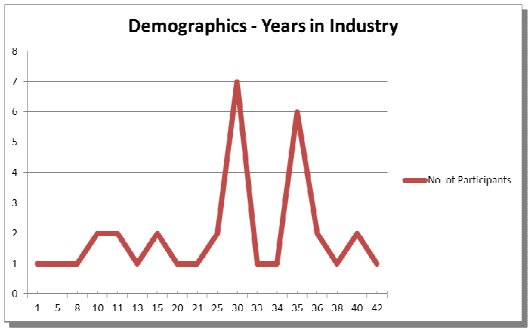


Figure 16 How many years have you worked in the Industry?

Figure 15 and Figure 16, provide an indication of the age and experience distribution of the industry participants. There can be seen to be a correllation between the age distribution and the years served in the industry, indicating that the longer serving industry participants are more willing to impart their knowledge, through participation in the this survey, as opposed to younger industry participants.

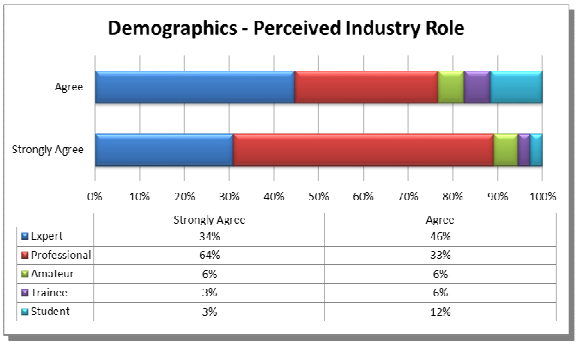


Figure 17 Which of the following do you consider yourself to be in your Industry?

This correllation is further strengthened, in Figure 17, where the industry role of Expert or Professional can be considered by the longer serving industry members. This provides a strong argument that the knowledge being shared within the Industry is trustworthy and relevant as it is originating from an expert source.

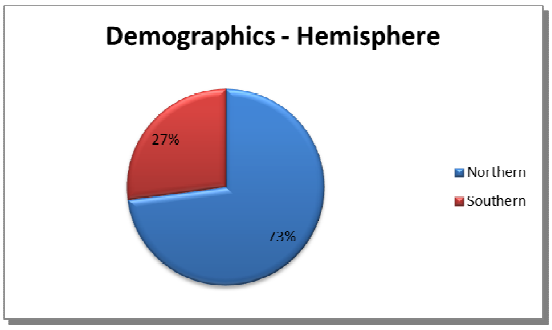


Figure 18 Geographical Hemisphere?

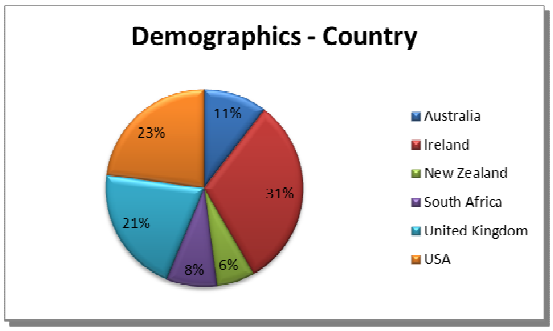


Figure 19 Country of Residence?

Figure 18, indicates that there is a larger distribution of survey participants from the Northern Hemisphere as opposed to the Southern Hemispheres. This distribution is confirmed by the results displayed in Figure 19, showing a higher proportion of Northern Hemisphere countries in the Pie Chart.

Figure 20 Survey participants plotted on world map, using logged IP addresses

The survey was distributed to 150 Stud Farms (46% of which were Southern Hemisphere), 50 Bloodstock Agents (42% of which were Southern Hemisphere) and 100 Trainers (20% of which were Southern Hemisphere), but the higher participation of Northern Hemisphere respondents may be an indication of the higher percentage of participation of Trainers as indicated in Figure 21.

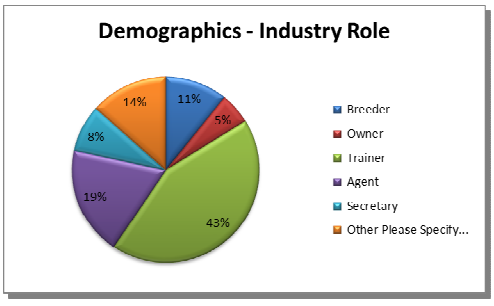


Figure 21 Please indicate your Industry Role?

The distribution of Industry Roles is portrayed in Figure 21, with a definite tendency towards Trainers. The originally distributed ratio of surveys was as follows; 3 : 2 : 1 in favour of Stud Farms : Trainers : Agents, respectively, see Figure 22.

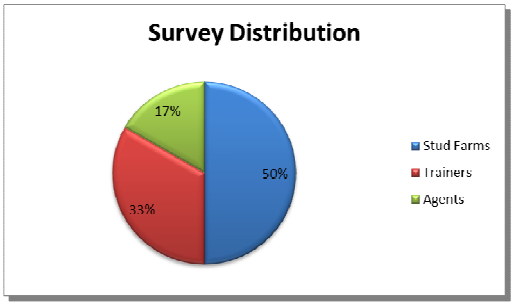


Figure 22 Original Survey Distribution

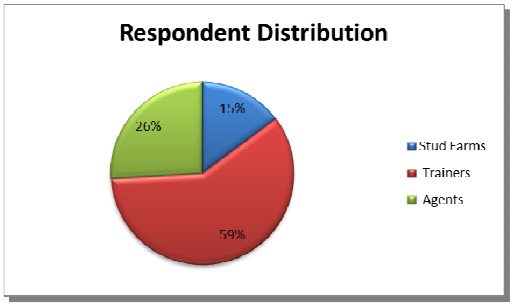


Figure 23 Adjusted Respondent Distribution

By adjusting the previous distribution of Industry Roles to remove ambiguous responses, such as Owner, Secretary, Other, a rotational shift is evident, see Figure 23, with the ratio of survey respondents at approximately 4 : 2 : 1 in favour of Trainers : Agents : Stud Farms respectively.

This may be viewed as an indication of the higher levels of technical ability on the part of Trainers in relation to the use of Blogs. It can also be perceived as an indicator that a higher percentage of Blogs and News sites available in the Thoroughbred Industry are predominantly based around the Horse Racing domain. This hypothesis will be further examined in the next chapter.

5.5 Blog Users Vs Bloggers

This section of questions has been used to gain insight into the technical proficiencies displayed by the industry users with respect to their use of online resources. A series of questions were presented to the survey participants to evaluate the use of Industry related News and Blog sites for information gathering. These were followed by a series of questions designed to determine the percentage of the survey participants who actively contribute to the body of online knowledge in the Industry, via blogs or other online resources.

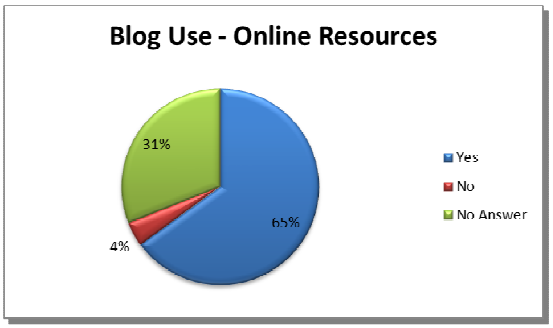


Figure 24 Do you use or read Online Resources for Industry Related News?

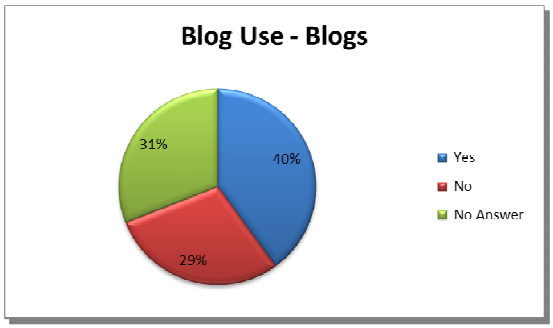


Figure 25 Do you use or read Industry Related Blogs?

As can be seen, Figure 24, shows that a high percentage of respondents use or read online resources for sources of Industry Related News for knowledge internalisation and information gathering. For the purpose of this research "Use" and "Read" were used as interchangeable terms, to allow for possible differences in terminology use by different survey participants. Figure 25 indicates that a slightly lower percentage of those respondents use or read Industry Related Blogs. While it can be inferred from the responses to the previous question, that the survey participants were aware of the difference between blogs and other types of online resources, some of the examples provided by the participants included:

- The Racingpost News Portal
- Bloodhorse.com News Portal
- Newsletters
- Equine Science Websites
- Weatherbys Stallion Register

Although the majority of these unprompted responses include News portals and other Information, as opposed to blogs, a further investigation was not carried out to determine which provided the greater sources of knowledge and information. This could form the basis of a future study, which is currently outside the scope of this research.

A further series of questions, were used to determine the numbers of survey participants that contribute to the body of online Industry Related Knowledge through personal blogging and/or other forms of web presence.

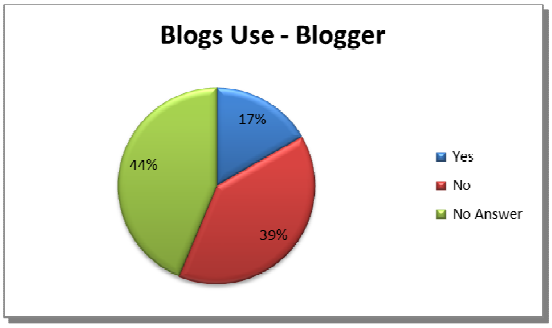


Figure 26 Do you maintain an Industry Blog?

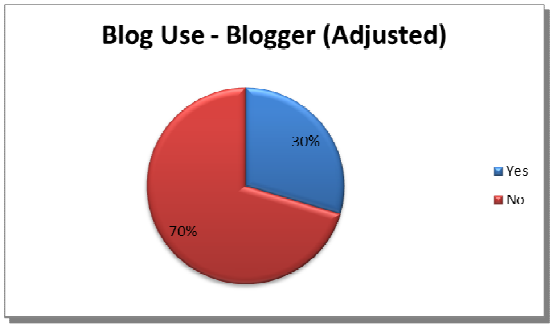


Figure 27 Adjusted Figures

A total of 44% of survey participants did not complete the series of questions related to their online presence, see Figure 26. The figures were accordingly adjusted, in Figure 27, to give a better representation, highlighting a 70-30 split between those that do not and those that do maintain an Industry Blog. This figure, almost a third of respondents, provides a representative indication of the contribution made by the survey participants to the body of online industry knowledge through their use of Blogs.

Even with newer social software technologies, such as Facebook and Twitter, being embraced by the Industry, conventional websites still provide the highest proportion of industry web presence, with 48% of survey respondents indicating that they maintain a website.

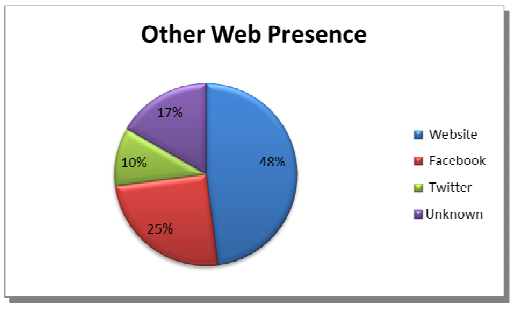


Figure 28 Do you maintain any other Online presence?

Although, conventional websites can be seen as the main conveyor of online presence, the adoption rate of Social Software of 35% (Facebook and Twitter) can be seen as a distinct move towards a more social web based industry.

5.6 *Purpose of Blog Use*

In order to establish the purpose of blog use by the survey participants a set of questions were presented in order to determine if industry blogs were being used for the purpose of knowledge sharing and internalisation.

Figure 29, provides an indication of the main purposes for reading industry related blogs. 26% of respondents indicate that they read industry blogs both for, interest and for Knowledge Internalisation. Reading blogs for educational purposes was cited by 20% of respondents, which can be directly related to knowledge internalisation. Revising the statistics, see Figure 30, therefore reveals that 46% of survey respondents read industry blogs for Knowledge Internalisation purposes.

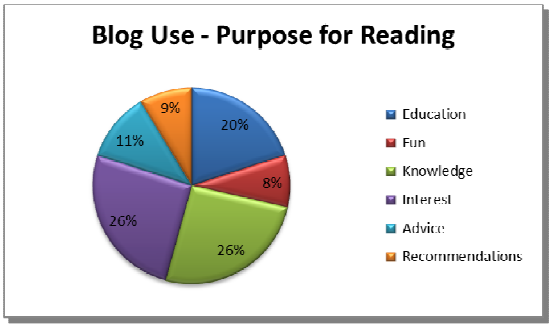


Figure 29 What is your main purpose for reading the Blogs?

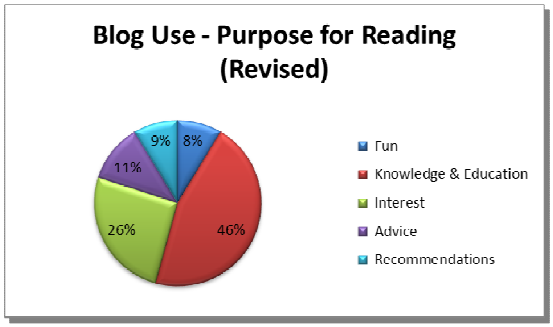


Figure 30 Purpose for Reading Blogs (Revised)

For the respondents that revealed that they maintain industry related blogs, Figure 32, indicates that 18% blog for the purpose of sharing knowledge. Similarly 18% revealed that the maintain a blog for sharing information. As with the previous analysis, a direct correlation can be drawn between knowledge and information sharing, with the figures being revised accordingly. The revised statistics, see Figure 32, imply that 37% of respondents blog for Knowledge Sharing purposes.

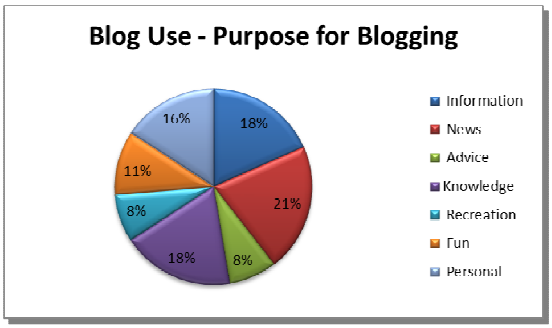


Figure 31 For what purpose do you maintain a Blog?

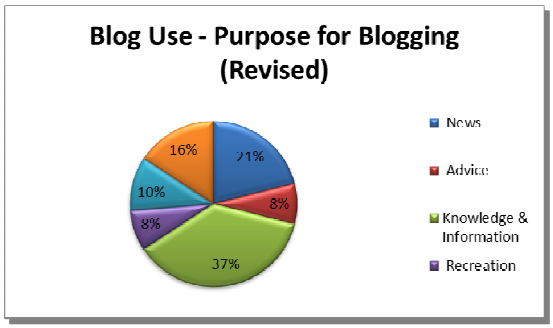


Figure 32 Purpose for Blogging (Revised)

The results of this series of questions provides evidence that the members of the Thoroughbred Industry use blogs for the purposes of Knowledge Internalisation and Sharing and thus offers further evidence that Weblogs can be leveraged as an enabler of Knowledge Sharing for the Industry as a whole.

5.7 Perception of Knowledge Sharing

Finally, in order to assess the Industry perception of blog use as a mechanism for Knowledge Sharing, the survey respondents were asked a number of questions to indicate how well blogs are at providing access to Industry Knowledge.

Figure 33 and Figure 34, provide strong evidence that members of the Thoroughbred Industry regard blogs as a good source of Industry Knowledge. 70% of survey respondents agree that blogs are a good source of knowledge, with 20% of that figure storngly agreeing with the proposition. A slightly higher, 75% of respondents, agree that blogs are a good source of knowledge for the industry, that can be used while travelling. Since a large portion of the roles in the Thoroughbred Industry involve travelling within and without native countries, this figure provides verification that blogs are an ideal mechanism for Industry Knowledge Internalisation.

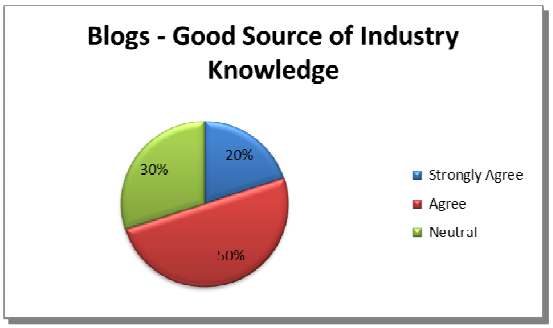


Figure 33 Blogs are a good source of Industry Knowledge

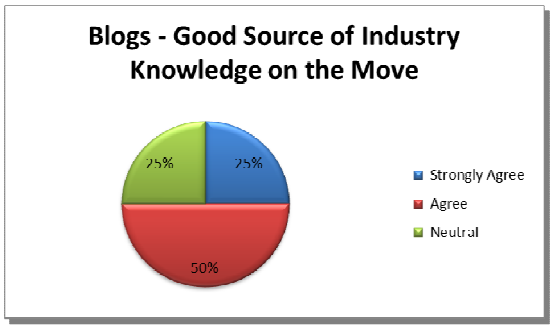


Figure 34 Blogs are a good source of Industry Knowledge when travelling

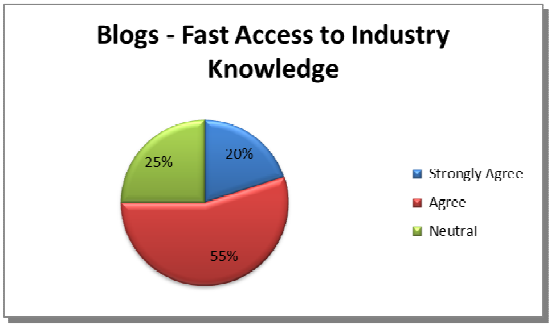


Figure 35 Blogs are a fast way of accessing Industry Knowledge

Comparably, 75% of survey respondents agree that blogs provide fast access to Industry Knowledge. This provides another key indicator that Weblogs can be leveraged as an enabler of Knowledge Sharing for the Industry, as knowledge can be accessed quickly and does not require extensive elicitation and acquistion from source experts.

5.8 Key Findings

- Longer serving industry participants are more willing to impart their knowledge
- The industry role of Expert or Professional can be considered by the longer serving industry members
- A definite tendency towards Knowledge Sharing by Trainers
- Higher percentage of Blogs and News sites available in the Thoroughbred Industry are predominantly based around the Horse Racing domain
- Websites still provide the highest proportion of industry web presence
- The 35% adoption rate of Social Software (Facebook and Twitter) can be seen as a distinct move towards a more social web based industry

- The members of the Thoroughbred Bloodstock Industry use blogs for the purposes of Knowledge Internalisation and Sharing
- Verification that blogs are an ideal mechanism for Industry Knowledge Internalisation
- Weblogs can be leveraged as an enabler of Knowledge Sharing for the Industry, providing fast access to Industry Knowledge.

5.9 Conclusions and Future Work

This chapter presented strong evidence indicating that blogs are a key enabler of Knowledge Sharing within the Thoroughbred Bloodstock Industry. The analysis of the demographics of the survey respondents provides evidence that the use of Industry Blogs supersedes previous global and cultural limitations. It also presented evidence that industry participants are more willing to impart their knowledge the longer they have served in the Industry, in the guise of experts and professionals.

A comparison of blog use and blog maintenance suggests that Industry blogs are used for knowledge internalisation as opposed to knowledge externalisation. This evidence provides a strong basis for the experiment to be carried out in the next chapter and points to a distinct shift towards a more social web based industry. The use of blogs as an enabler of Knowledge Sharing was also confirmed presenting evidence that there has been a definite shift in knowledge management and sharing practices in the Thoroughbred Bloodstock Industry, from Socialisation to Externalisation.

Overall the results of the survey provide strong evidence that Weblogs can be leveraged as an enabler of Knowledge Sharing for the Industry. The next chapter will seek to analyse how blog and RSS technologies can be exploited in an attempt to promote the industry from a social to a socially semantic web based industry.

6 RSS AND CONTENT ANALYSIS

6.1 Introduction

Until recently the Thoroughbred Industry has been a predominantly informal social industry sharing knowledge via "word of mouth", "nods and winks", "handshakes", etc. This socialisation of knowledge has led to an immense accumulation of Tacit Knowledge within the Industry. With the advent of Web 2.0 and Social Software, industry experts, professionals and enthusiasts alike have found a new medium for conveying their knowledge, however the methods used to organise the knowledge exposed through "knowledge conversation" is still extremely informal. Although this informality provides a quick and easy method for conveying knowledge, it does not provide an adequate method for organising the knowledge, in order to maximise its usefulness to the community as a whole. The Thoroughbred Industry tends to use a descriptive terminology, concentrating on the familiarity of the community with names of Horses, Trainers, Breeders, Studs, Racecourses, etc. As a result a formal method of knowledge organisation does not exist, but the development of a basic domain specific ontology, derived from the informal folksonomy within the community, could be seen as a move towards a more formal or Social Semantic Industry.

This Chapter will assess the suitability of RSS for the conveyance of organised domain knowledge and use feed content analysis to propose a process of key term identification to develop an industry folksonomy. A basic domain-specific ontology will be developed using emergent folksonomy terms and a Hybrid Tagging System proposed to bootstrap the interaction between an industry folksonomy and domain-specific ontology. Finally a discussion on possible future work will be presented, outlining additional experimental steps that could be taken to improve the results of the Industry Folksonomy.

6.2 Initial Review

An initial review of a cross-section of Industry Blogs, the Tags employed by the bloggers and their syndication feeds, was carried out. This review was conducted in order to assess the familiarity of the community with the Web 2.0 technologies

available for knowledge sharing and their proper and consistent usage. (A complete list of the Industry Blogs reviewed and used in this chapter can be found in Appendix B).

6.2.1 Industry Representation of Blogs

In total 17 individual Blogs or News related sites were reviewed and assessed for their applicability for inclusion in this experiment, identified in Table 5. The sites reviewed consisted of Northern and Southern Hemisphere, Thoroughbred Racing and Breeding, and Expert and Amateur Blogs and News related portals.

Blog / News Site	**Hemisphere**	**Industry Domain**	**Type of Contributor**
Arqana Sales News	Northern	Sales/Breeding	Sales Organisation
BBC World Edition	Northern/Southern	Racing	Media
BHA Handicapper	Northern	Racing	Prof. Organisation
BHA News	Northern	Racing	Prof. Organisation
BHA Xtra	Northern	Racing	Prof. Organisation
Calvados Bloodstock	Northern	Breeding/Racing	Industry Professional
Irishracing.com	Northern	Breeding/Racing	Prof. Organisation
Paulick Report	Northern/Southern	Breeding/Racing	Industry Professional
Racingpost Bloodstock	Northern/Southern	Breeding	Media
Racingpost Racing	Northern	Racing	Media
Jessica Chapel / Railbird	Northern	Breeding/Racing	Amateur Blogger
Saratoga News	Northern	Racing	Industry Professional
Summerhill Blog	Southern/Northern	Breeding/Sales/Racing	Industry Expert
TBA	Northern	Breeding	Prof. Organisation
TDN	Northern/Southern	Breeding/Sales/Racing	Media
Thoroughbred Bloggers Alliance	Northern/Southern	Breeding/Sales/Racing	Expert / Amateur
Thoroughbred Horse Racing	Northern	Racing	Industry Professional

Table 5 Reviewed Blogs and News Sites

Each of the selected Blogs provide regular feed updates, either daily, weekly or monthly to keep their audience appraised of up-to-date information and knowledge relevant to the blogs Industry Domain. The types of contributors identified with each of the blogs provides inference that the intended audiences will be diverse in nature, ranging from Industry Experts to Industry Amateurs and also including non-industry related spectators and enthusiasts.

The chosen review sites serve as a representative cross-section of easily identifiable and well established industry sources, including the BBC (British Broadcasting Company), the Racing Post (leading British daily horse racing and sports newspaper), the BHA (British Horseracing Authority) and some newer sites with increasing popularity within the community, e.g. Summerhill Blog (South African Breeding Champions for Sixth Consecutive Year).

6.2.2 Lack of Exploitable Feed Tags

An examination of the feed structures from the chosen Blogs and News Sites was effected in order to determine the usefulness of the categorisation or tagging of the feed articles. The XML files were first reviewed using Syntext Serna (a Multiplatform WYSIWYG XML Authoring application) to assess the well-formedness of the XML documents and the existence of categories and/or tags. The feeds were also consumed using, RSSOwl, a popular Eclipse based RSS Reader to ensure that the feeds could were compliant with feed standards.

While all the feeds were standards compliant and could be consumed using RSSOwl, it was found that there was varying degrees of tagging consistency among the tested feeds. In some cases no categories or tags were included in the feeds to help organise or identify the feed article. In the remainder of instances the categories or tags used were found to be too low level, or descriptive, to be extracted and used for industry concepts. The lack of suitable tagging meant that a statistical analysis of Industry Terms useful for production of a basic folksonomy would not be possible.

The lack of high level, or terminological, tagging within the feeds was attributed to the predominantly tacit to tacit nature of knowledge sharing extant in the industry. It can thus be suggested that the possibility that the domain is without a history of explicit knowledge sharing that tags are still representative of the expectant tacit knowledge of the readers, i.e. Horse Names, Trainer Names, Race Names, Jockeys, etc., see Figure 36 and Figure 37.

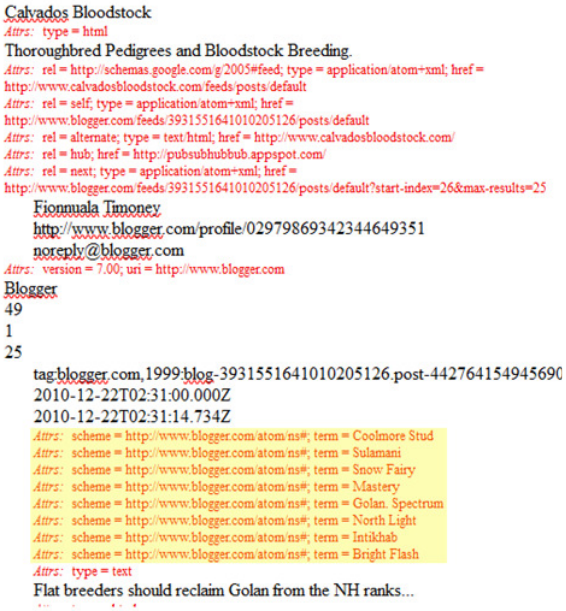


Figure 36 Sample of descriptive categories/tags used in Calvados Bloodstock ATOM feed

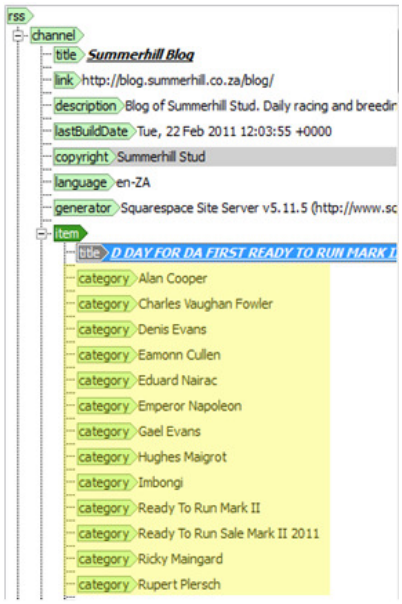


Figure 37 Sample of descriptive categories/tags used in Summerhill RSS feed

Following this initial review stage and the lack of exploitable tags within the surveyed feeds, it was decided to change the focus of the experiment and instead perform an analysis of feed article content in the hopes that this would produce a better distribution of Industry Terms that could be used for production of a basic folksonomy.

6.3 Feed Content Analysis

In computational linguistics, a frequency list is a list of, where the frequency usually means the number of occurrences of the word in a given corpus, text or content. These word frequency lists are usually sorted by the frequency of occurrence of the words. In qualitative research, content analysis, is commonly regarded as simply doing a word frequency count, with the assumption made that the most frequently mentioned words are those that are attributed the greatest importance. While this may be true in some cases, there are several counterpoints to consider when using simple word frequency counts to make inferences about matters of importance.

Weber (1990) indicates that synonyms may be used throughout a text for stylistic reasons which can lead to the underestimation of a word or concepts importance during automated content analysis. Additionally, each word may not represent a category

equally well, and with a lack of well-developed weighting procedures to identify this, using word counts requires the researcher to be aware of this limitation.

The decision to perform an analysis of the feed content required acceptance of a number of limitations, namely the lack of a suitable weighting system for keyword importance to apply to the content analysis. It was felt that this limitation was acceptable as the analysis was being conducted in order to determine basic Industry concepts and not necessarily their importance. Additionally, simple word extraction is based on word frequency while complex ones use statistical techniques or linguistic techniques supported by domain specific ontologies. The lack of a suitable domain specific ontology was also a determining factor in the decision to use a more simplistic analytical approach. It was hoped that the identified Industry concepts would provide a starting point for proposal of a domain specific ontology.

The main concession that had to be made was the use of a 3rd party RSS reader. The development of a ColdFusion based Feed Reader was planned and started, see Figure 38, with the intention of consuming, storing and analysing the Industry feeds. An initial feed reader was developed which allowed feeds be consumed, but time restrictions and lack of required knowledge meant that complete development of a storage facility, text parsing (for removal of inline HTML tags) and lexical analysis features could not be completed. As a result it was decided to use other freely available tools which had been designed with each of these purposes in mind.

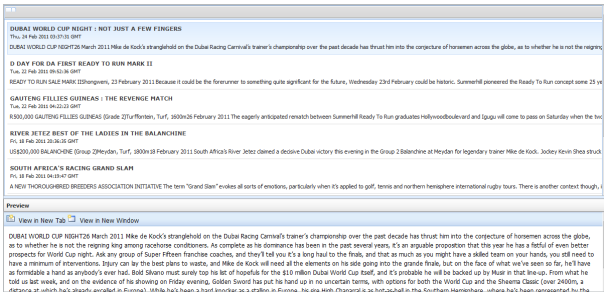


Figure 38 Sample screenshot of partially developed ColdFusion Feed Reader

The main experiment was conducted in four phases:

1. Critique of RSS Readers and consumption of RSS Feeds - The first phase was to identify the most suitable RSS reader to be used to consume the cross-sectional representation of Industry Blog feeds identified in Section 6.2.1.

2. Preparation for Analysis - The second phase, involved manual preparation of the feed content for use in an online Text Analyser. Feed contents were exported to HTML and then copied to Notepad (to remove HTML tags) in preparation for word frequency analysis.
3. Analysis Tool - The third phase consisted of analysis of each Notepad text file using the online Text Analyser at UsingEnglish.com to produce detailed statistics of common words and their frequency of use in the texts. (http://www.usingenglish.com/members/text-analysis/)
4. Analysis Results - The final phase was to extract the statistical data from the Text Analyser and store the results in an MS SQL database. This would facilitate further analysis and result charting in MS Excel.

6.3.1 Selection of RSS Readers

During this phase of the experiment a number of free, desktop and web-based, RSS Readers were assessed and tested in order to determine which software would provide the facilities required. These included:

- RSS Owl
- iGoogle
- Google Reader
- Blogbridge

All four of the tested readers performed equally well with respect to consuming the Blog and News Site feeds. All four readers displayed the feed article correctly formatted, including inline images, text formatting, web links, etc. Further testing exposed some deficiencies, which would prove critical deciding factors in the selection of the accepted RSS reader.

RSSOwl

The application selected for the consumption of the RSS Feeds, was RSSOwl, a news aggregator for RSS and Atom News feeds, written in Java, and built on the Eclipse Rich Client Platform. It proved to be the most robust and feature packed of the four tested readers, providing full support for RSS & RDF versions 0.91, 0.92, 1.0, 2.0 and

support for Atom Syndication Format version 1.0. RSSOwl's ability to import the feed article categories, provided search and sort capabilities based on the categories. However, the lack of feed content analysis functionality meant that some additional factor would be required for consuming and exposing the feed articles.

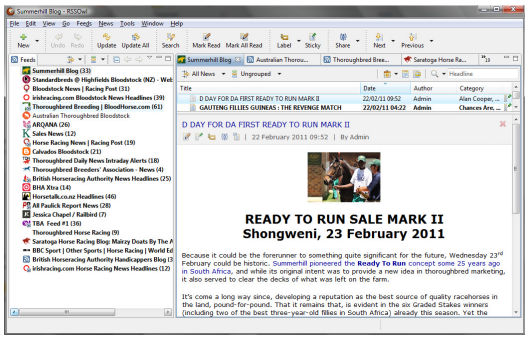


Figure 39 Example of Summerhill Blog feed consumption in RSSOwl

RSSOwl's ability to generate PDF, RTF, HTML and XML documents from any news including aggregations provided the determining factor. Exposing the feed articles in this way would make them more accessible for the required content analysis.

iGoogle

iGoogle, a service of Google, is a customisable Ajax-based startpage or personal web portal, whose features can be extended with the capability to add web feeds and Google Gadgets. While an unlimited number of RSS feeds could be added and consumed by iGoogle, it provided little in the way of functionality other than displaying the article summaries and linking to the original Blog article via its embedded permalink. It provided no means of organising the feeds based on included categories or tags, as these did not form part of the feed consumption. No analysis features are included in iGoogle which might allow the user to analyse the text of a feed article. Additionally, no facility to export articles or their details is available to the user.

Figure 40 Example of Summerhill Blog feed consumption in iGoogle

Google Reader

Google Reader, also a service provided by Google, is a Web-based aggregator, capable of reading Atom and RSS feeds. Feed articles can be organised with labels, as well as users being able to create "Starred Items" for easy access, in a similar fashion to Gmail. However, it was found that again, the categories or tags included in the RSS feeds were not used during feed consumption. As with iGoogle, Google Reader, provides no methods for analysis of feeds or exporting feed article contents for further analysis.

Figure 41 Example of Summerhill Blog feed consumption in Google Reader

BlogBridge

BlogBridge is an open source Java-based feed aggregator, which is aimed for those people who are subscribed to many feeds. As with the previous readers it was capable of consuming the RSS feeds and displaying their content in a well formatted fashion. BlogBridge differentiated itself as it also imported the feed article categories during the RSS feed consumption, but these were found to provide no organisational functionality and served only as display or reference.

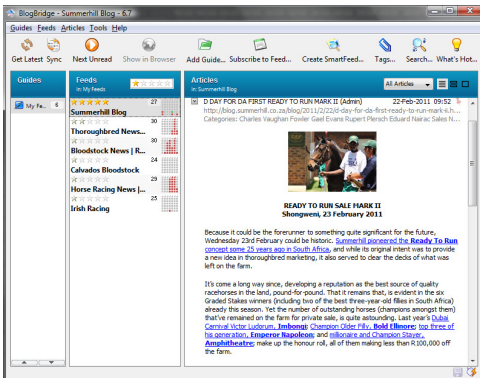


Figure 42 Example of Summerhill Blog feed consumption in BlogBridge

Although BlogBridge displayed a good representation of the content of the feeds, it did not provide any analysis tools or means of exporting the content of the feed for analysis with an external tool.

	iGoogle	Google Reader	BlogBridge	RSSOwl
Consume RSS	✓	✓	✓	✓
Consume ATOM	✓	✓	✓	✓
Import Tags				✓
Support Tagging		✓	✓	✓
Support Sharing		✓	✓	✓
Tag Organisation		✓	✓	✓
Export Facility				✓
Built in Analysis				

Table 6 Feature comparison matrix for tested RSS Readers

Table 6, provides a feature comparison matrix for the tested RSS Readers. Each of the RSS Readers were tested based on these features and as previously indicated RSSOwl's superior feature set identified it as the RSS Reader most suited for use in this project.

6.3.2 Preparation for Analysis

The feed article contents for each of the selected Industry Blogs were exported from RSS Owl and saved in individual HTML formatted files. Each of these HTML files were in turn opened and viewed in Microsoft Internet Explorer in order to extract the text.

The text extraction was performed manually by selecting the entire HTML webpage in Internet Explorer and the contents copied to Notepad. The process of copying the text to Notepad was preceded by an intermediary stage which involved performing a word and character count of the text in Microsoft Word. This was required as the online Text Analysis Tool used in the third phase had an upper limit of 100,000 characters, which could be analysed at one time. Text that was found to contain more than 100,000 characters was split between two separate text files in order to facilitate the upper limit of the Text Analysis Tool.

The removal of underlying HTML tags was a key factor in improving the lexical analysis results, by excluding as many tags and characters that could contribute to noisy results as possible. Since Notepad supports only very basic formatting, by subsequently copying the counted text into Notepad, special formatting, including Images and underlying HTML tags, were automatically removed leaving pure text, acceptable for further analysis.

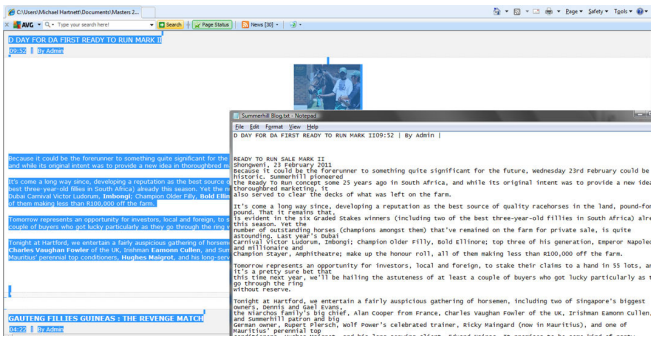


Figure 43 Exported feed articles being copied from Internet Explorer to Notepad

No further text manipulation was required during this phase as the text files now represented the parsed and cleaned content to be used for lexical. (Complete HTML and plain Text files used for analysis are included on CD at the back of this Dissertation).

6.3.3 Analysis Tool

In order to perform a qualitative content analysis, of the RSS feeds, a method for counting the frequency of the words was required. Although simple frequency counts of the words will help to identify common words used within the feeds, further analysis would be required to identify those words that have domain relevance and describe industry concepts.

UsingEnglish.com is a specialist English language site, with a wide range of resources for learners and teachers of English, and has been running since the beginning of 2002. Different varieties of English are used; with contributors from the United States, Canada, Pakistan and non-native speaking countries, but much of the site uses British English as it was set up in the UK.

UsingEnglish.com also provides a large collection of English Language tools and resources for students, teachers, learners and academics. The free text analysis tool provided on the website can analyse text content resulting in statistics about the text including:

- Word Count
- Word Distribution
- Unique Words
- Number of Sentences
- Average Words per Sentence
- Lexical Density
- Gunning Fog Readability Index

The decision to utilise the free text analysis tool on the UsingEnglish.com website was based a number of determining factors. First, it was hoped that since the website contributors were from a global background that this would benefit the text analysis as the feed contents were from a global representation. Secondly, the text analyser provided a complete breakdown of word frequency counts which could be stored locally for further analysis (in Section 6.3.4). Finally, the word distribution statistics delivered for each of the analysed texts would help establish the optimal starting point for extracting the common industry terms.

The basic text analysis tool was limited to analysing only 10,000 characters, but this limit was increased to 100,000 characters per block of text, upon registration with the site. Even with the increased limit, some feed content had to be split into two sections for analysis. This did not affect the outcome of the final results as the sum of word

frequency for the two sections would equal the total word frequency if the text was analysed as a single block.

There was also a limit on the total number of saved text analyses that could be stored even for registered site users. This limit was 20 analyses, but with 3 feeds exceeding the 100,000 character limit and needing to be split, only 17 feeds could be analysed and stored for further statistical analysis. The following section describes the processes involved in procuring frequencies, storing, cleansing, processing and further analysing the analysed feed contents.

6.3.4 Analysis Results

Following online text analysis, the statistical word frequency results were copied and formatted in MS Excel. An MS Excel spreadsheet (included on the accompanying CD) was created to store the word frequencies and add additional information including the feed source, which was used to help contrast terms from different areas of the Industry. The following is a 6 step '*algorithm*' used during the analysis of the extracted RSS feed content. Each of the 6 steps is identified and explained as follows:

1. Compile Stop Words

Stop words are words which are filtered out prior to, or after, processing of natural language data (text). The list of stop words is controlled by human input and not automated. There is not one definite list of stop words which all tools use, so a number of stop word lists were combined for data cleansing during this phase of the experiment.

To this combined stop word list was added a number of extra word sets. These included Days of the Week, Months of the Year, etc. and over 1,900 of the most common English Christian Names sourced from The Christian Names Dictionary (http://www.babynames.org.uk/christian-names-dictionary.htm). The additional stop words and names were included in an attempt to remove extra words which would attribute to noisy or outliers in the results.

Once all were compiled in the MS Excel spreadsheet, the next step involved importing the word frequency and stop word lists into MS SQL database, where formal queries could be run against the data in order to produce concise lists of terms or concept words which are frequently used in the Thoroughbred Industry.

2. Identify Optimal Starting Point

Many optimisation algorithms need to start from an optimal starting point in order to perform feasible analysis. In the context of this research, identifying the optimal starting point will establish the word length which provides the lowest level of applicable domain relevance, i.e. the minimum length of a word that can describe an Industry concept.

In order to determine the optimal starting point for the analysis the word distribution statistics from the text analyses were examined and in each case the most frequent word distribution was found to be amongst 3 letter words. This would help determine the minimum word length for the most frequently used words across all analyses and with this information, words with less than 3 letters were discarded. Although this may seem quite a low common denominator, industry experience has helped guide this decision also as quite a number of important Industry concepts and terms fall into the 3 lettered category including; ‘dam’, ‘win’ ‘cup’, ‘lot’, etc.

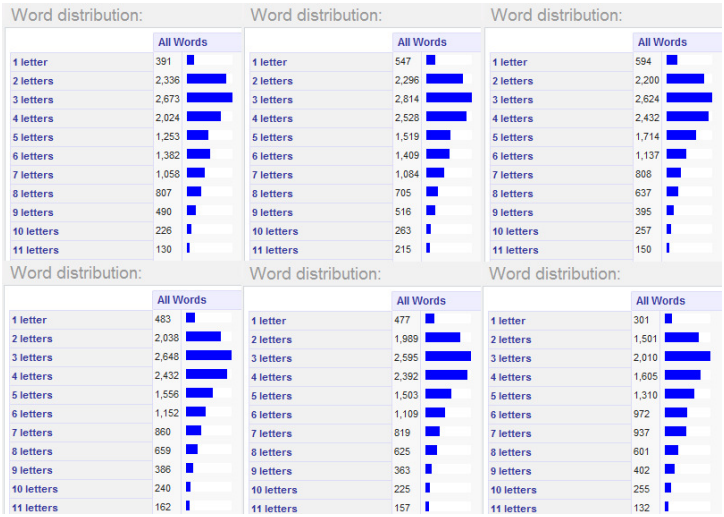


Figure 44 Sample word distribution from text analysis of the six largest feed contents

3. Implement SQL Queries and Views

Based on the compiled stop word lists and the determination of the optimal starting point for the data analysis, a number of SQL queries were developed which were used to create views on the underlying imported data. These queries would provide a cleansed and significant view of the words, their frequencies and common sources for further examination. These views were also be exposed and accessed via MS Excel to provide charting capabilities to provide a visual representation of the resultant findings.

```
SELECT    TOP (50) word, SUM(count) AS Frequency, COUNT(word) AS Feeds
FROM      dbo.WordAnalysis
WHERE     (length > 2) AND (word NOT IN
                    (SELECT DISTINCT StopWord
                     FROM       dbo.StopWords)) AND (word NOT IN
                    (SELECT DISTINCT Name
                     FROM       dbo.StopNames))
GROUP BY word
ORDER BY Frequency DESC
```

Figure 45 SQL query used to retrieve the Top 50 most frequently used words based on frequency

```
SELECT    TOP (50) word, SUM(count) AS Frequency, COUNT(word) AS Feeds
FROM      dbo.WordAnalysis
WHERE     (length > 2) AND (word NOT IN
                    (SELECT DISTINCT StopWord
                     FROM       dbo.StopWords)) AND (word NOT IN
                    (SELECT DISTINCT Name
                     FROM       dbo.StopNames))
GROUP BY word
ORDER BY Feeds DESC
```

Figure 46 SQL query used to retrieve Top 50 most frequently used words based of no. of feeds

```
SELECT    TOP (50) word, SUM(count) AS Frequency, COUNT(word) AS Feeds
FROM      dbo.WordAnalysis
WHERE     (length > 2) AND (word NOT IN
                    (SELECT DISTINCT StopWord
                     FROM       dbo.StopWords)) AND (word NOT IN
                    (SELECT DISTINCT Name
                     FROM       dbo.StopNames)) AND (source IN
                    (SELECT    FeedSource
                     FROM       dbo.Source
                     WHERE     (FeedType LIKE '%R%')))
GROUP BY word
ORDER BY Frequency DESC
```

Figure 47 SQL query to retrieve Top 50 most frequently used words from Racing specific feed sources

```
SELECT    TOP (50) word, SUM(count) AS Frequency, COUNT(word) AS Feeds
FROM      dbo.WordAnalysis
WHERE     (length > 2) AND (word NOT IN
                    (SELECT DISTINCT StopWord
                     FROM       dbo.StopWords)) AND (word NOT IN
                    (SELECT DISTINCT Name
                     FROM       dbo.StopNames)) AND (source IN
                    (SELECT    FeedSource
                     FROM       dbo.Source
                     WHERE     (FeedType IN ('B', 'SB'))))
GROUP BY word
ORDER BY Frequency DESC
```

Figure 48 SQL query to retrieve Top 50 most frequently used words from Breeding specific feed sources

4. Additional data cleansing and examination

Initial review of the data presented by the SQL views identified some outliers, or noisy results in the form of word pluralisation and contraction. In some cases the inclusion of these word forms meant that significant words or terms were being excluded. Examples include:

Examples of plural transformations		
horses	→	horse
trainers	→	trainer
colts	→	colt
fillies	→	filly

Examples of word contraction transformations		
races	→	race
season's	→	season
breeders'	→	breeder
sire's	→	sire

Table 7 Examples of plurals and word contractions that were reviewed to remove outliers and noisy results

Removal of plurals and contractions was performed manually by querying the database and reviewing the word contexts. Straightforward plurals and contractions were replaced with the singular, e.g. 'races' replaced by 'race'. This form of data cleansing served to enforce the word frequency, which would improve the developing folksonomy, and enhance the concepts and classes that could be used to develop a domain specific ontology.

Other variants of words, including past and present participles, were assessed and dealt with according to the context and the word. In most cases words found with these properties were left unmodified as they were found to be contextually relevant as usable Industry terms and concepts. Examples included past and present participles of the word 'race', both 'raced' and 'racing' describe valid terms and concepts in use within the Thoroughbred Industry. 'Raced' is used as a classification class for racehorses, to indicate if the horse has ever run in a race, i.e. raced or unraced. Likewise, the word 'Racing' is a commonly shortened construct of the word 'horseracing', used to classify the area of the Industry.

During the process of word participle examination, it was found that a common industry classification was missing from the returned queries of cleansed words. "Going" is a common stop word but is also a valid conceptual term in the Thoroughbred Industry used to describe the condition of a racecourse, or racetrack, for racing, i.e. 'firm', 'soft', 'good to firm', etc. The word was removed from the stop word list and subsequently found to be one of the top 25 most frequently used words in the analyses.

This presents an interesting finding and highlights a possible requirement for customised stop word lists applicable to industry specific domains. It also implies that

careful examination must be carried out when removing plurals, contractions and word participles to ensure that relevant industry concepts are not removed, leading to incomplete domain modelling.

5. Further Word Analysis

Further word variant analysis was conducted on both the experimental dataset and through knowledge of the domain which highlighted words that fall into distinct categories such as Hyponyms, Synonyms, Acronyms, etc.

Category	Meaning	Findings
Hyponyms	Word or phrase whose semantic range is included within that of another word.	Sire Stallion (both are hyponyms of Horse)
Antonyms	Words that lie in an inherently incompatible binary relationship as in opposite pairs.	Sire : Dam Colt : Filly
Acronyms	Abbreviations that are formed using the initial components in a phrase or name. (Acronyms in the Thoroughbred Industry tend to abbreviate Organisation Names or Veterinary Treatments).	BHA (British Horseracing Authority), TBA (Thoroughbred Breeders Association)
Quasi-Synonyms	Term designating the same concept as another, but which is not interchangeable in all contexts.	Jockey Rider
Monosemy	Words having a single meaning.	Owner Breeder
Abbreviations	A shortened for of a word or phrase. (Commonly used for Horse Colour)	Bay-Brown → bbr Black → bl
Synonyms	Different words with identical or at least similar meaning.	Racecourse (Europe) Racetrack (USA) (shows distinction between different continents)
Polysemy	Word or phrase with multiple, related meanings.	Maiden, in racing: a horse that has not won a race, in breeding: a female horse that has never been bred.
Hypernyms	A word, usually broad in meaning, which other more specific words are encompassed	**Horse** denotes: Stallion (Male horse)

	by.	Mare (Female horse) Colt (Young male horse) Filly (Young female horse)

Table 8 Further Word Analysis

Another interesting deviation in terms is the use of Sire, Grandsire and Damsire. Each word refers to a male parent in a Horse hierarchical pedigree or family tree, but each have distinctive references. The sire is the direct male parent of a horse, the grandsire is the paternal male parent and the damsire is the maternal male parent. These three words could be considered both Hyponyms and Hypernyms as their lexical relationships tend to cross the boundaries of both categories. Sire, grandsire and damsire are all hyponyms of Horse, while sire could be regarded as a Hypernym of both grandsire and damsire.

While this analysis of key word variants was informative in nature, it was not considered to be a relevant process required for the data cleansing, as most of the words or terms discussed are integral to forming an industry folksonomy.

6. Data Analysis

Subsequent to the additional data cleansing and word analysis the following results were found and reviewed to determine if a valid Industry Folksonomy was emerging. Figure 49, plots the Top 50 Common Industry Words by their frequency of use as found during the feed text analysis.

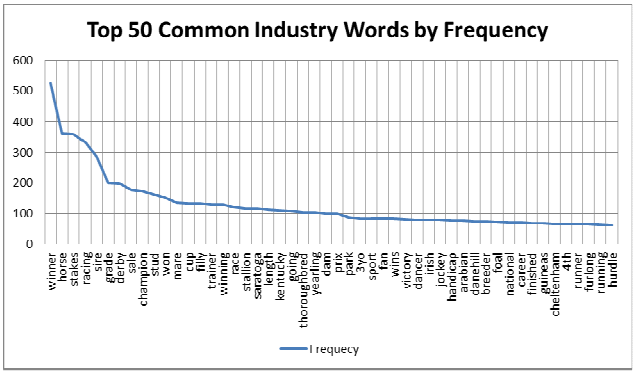


Figure 49 Top 50 Common Industry Words by Frequency

It is evident from this initial plot of frequently used industry words that there is an immediate tendancy towards the racing area of the Thoroughbred Industry. This is apparent from initial examination of the plotted terms, with the key concept of 'racing'

appearing in the top 5 results. Also, 'horse', the key class for the entire domain has emerged as the second most frequetly used term, providing early evidence that the content analysis is producing a viable folksonomy.

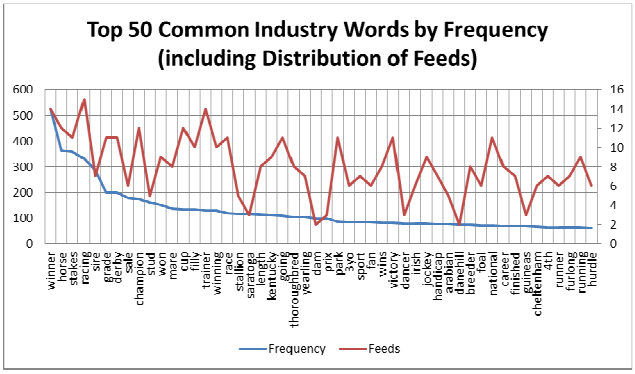


Figure 50 Top 50 Common Industry Words by Frequency (including Distribution of Feeds)

Figure 50, plots the same Top 50 Common Industry Words by their frequency, with inclusion of the Distribution of feeds that use the words. The visualisation of the feed distribution against the word frequency, shows definite concentrations of word usage in smaller distributions, e.g. usage of 'dam' is limited to approximately 2 feeds, but has a high frequency, or concentration, of use appearing 25th of the top 50. Conversely, evidence shows that the use of 'trainer' a term predominantly associated with the racing area of the Industry, is ranked 15th of the top 50, but is used in approximately 14 feeds.

Figure 51, plots the Top 50 Common Industry Words by the feed distribution of their use. This displays and even higher tendancy towards the racing area of the Thoroughbred Industry, with almost all of the Top 50 being terms specific to Horseracing.

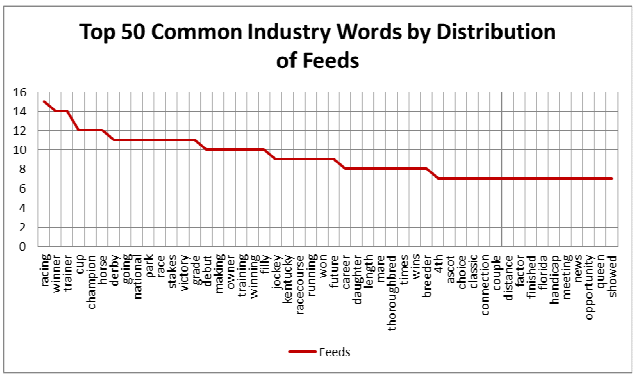


Figure 51 Top 50 Common Industry Words by Distribution of Feeds

Figure 52, displays the word frequency associated with the previously plotted Top 50 Common Industry Words by feed distribution. Again, this plot displays a trend that some of the more common, or important, Industry words are identified by their concentration of distribution as opposed to their individual frequency.

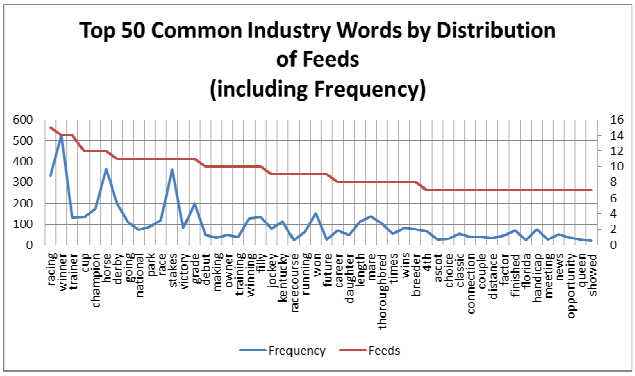


Figure 52 Top 50 Common Industry Words by Distribution of Feeds (including Frequency)

By way of domain separation individual views of the data were examined based on the two main Industry areas, Racing and Breeding. Figure 53, plots the Top 50 most commonly used Industry Words sourced from racing feeds in the source data.

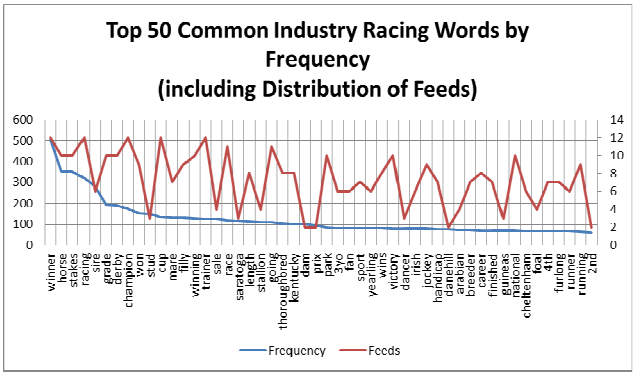


Figure 53 Top 50 Common Industry Racing Words by Frequency (including Distribution of Feeds)

Figure 54, plots the Top 50 most commonly used Industry Words sourced from breeding and sales related feeds in the source data. While Figure 53, shows only a 8% crossover of word usage between the two areas (2 in the top 25, i.e. ‘sire’, ‘dam’), Figure 54 displays a more than doubled increase, or 20%, crossover of word usage (5 in the top 25, i.e. ‘winner’, ‘racing’, ‘stakes’, ‘derby’, ‘grade’).

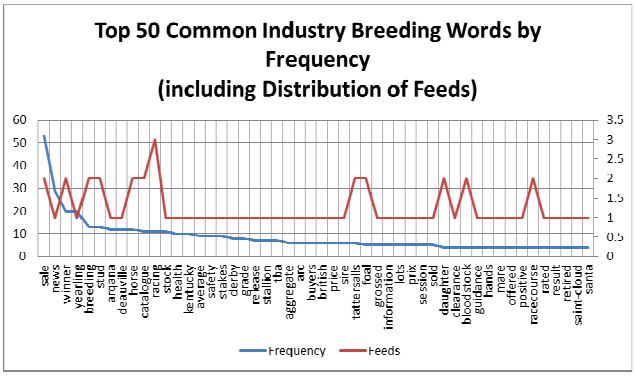
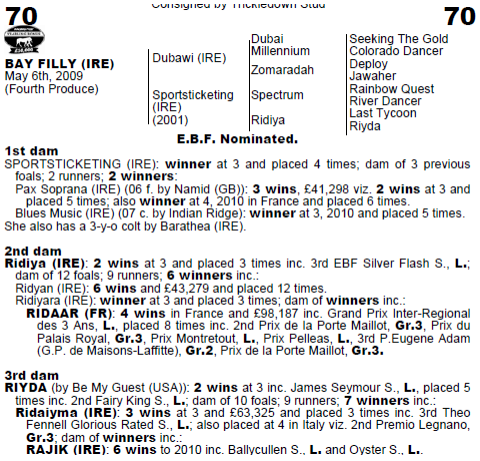


Figure 54 Top 50 Common Industry Breeding Words by Frequency (including Distribution of Feeds)

The crossover of word conventions between the racing and breeding domains is to be expected as the thoroughbred breeding sector is driven by an underlying philosophy, and the basis for improvement of the breed, of breeding superior mares to stallions that possess desirable attributes. While this philosophy does not guarantee that breeding a champion mare to champion sire will produce a future winner, it is based on extensive research and analysis of the racing qualities of the ancestry of the female line, crossed with the proven results of the stallion. This forms the basis of the domain crossover and is particularly evident when examining catalogue pages for Horses being sold.



70 **70**

BAY FILLY (IRE)
May 6th, 2009
(Fourth Produce)

Dubawi (IRE) — Dubai Millennium; Zomaradah
Sportsticketing (IRE) (2001) — Spectrum; Ridiya

Seeking The Gold
Colorado Dancer
Deploy
Jawaher
Rainbow Quest
River Dancer
Last Tycoon
Riyda

E.B.F. Nominated.

1st dam
SPORTSTICKETING (IRE): **winner** at 3 and placed 4 times; dam of 3 previous foals; 2 runners; **2 winners**:
Pax Soprana (IRE) (06 f. by Namid (GB)): **3 wins**, £41,298 viz. **2 wins** at 3 and placed 5 times; also **winner** at 4, 2010 in France and placed 6 times.
Blues Music (IRE) (07 c. by Indian Ridge): **winner** at 3, 2010 and placed 5 times.
She also has a 3-y-o colt by Barathea (IRE).

2nd dam
Ridiya (IRE): **2 wins** at 3 and placed 3 times inc. 3rd EBF Silver Flash S., **L.**; dam of 12 foals; 9 runners; **6 winners** inc.:
Ridyan (IRE): **6 wins** and £43,279 and placed 12 times.
Ridiyara (IRE): **winner** at 3 and placed 3 times; dam of **winners** inc.:
RIDAAR (FR): **4 wins** in France and £98,187 inc. Grand Prix Inter-Regional des 3 Ans, **L.**, placed 8 times inc. 2nd Prix de la Porte Maillot, **Gr.3**, Prix du Palais Royal, **Gr.3**, Prix Montretout, **L.**, Prix Pelleas, **L.**, 3rd P.Eugene Adam (G.P. de Maisons-Laffitte), **Gr.2**, Prix de la Porte Maillot, **Gr.3.**

3rd dam
RIYDA (by Be My Guest (USA)): **2 wins** at 3 inc. James Seymour S., **L.**, placed 5 times inc. 2nd Fairy King S., **L.**; dam of 10 foals; 9 runners; **7 winners** inc.:
Ridaiyma (IRE): **3 wins** at 3 and £63,325 and placed 3 times inc. 3rd Theo Fennell Glorious Rated S., **L.**; also placed at 4 in Italy viz. 2nd Premio Legnano, **Gr.3**; dam of **winners** inc.:
RAJIK (IRE): **6 wins** to 2010 inc. Ballycullen S., **L.** and Oyster S., **L.**.

Figure 55 Sample Goffs sales catalogue page showing analysis of Maternal lineage

In each of the inspections and plotting of Word Frequencies, the resultant charts and figures tend toward agreement with the theory of Power Law distributions. A Power Law is a special kind of mathematical relationship between two quantities in which the frequency of an event varies as a power of some attribute of that event. In the context of this experiment it is being used to demonstrate ranking of popularity or frequency of

industry words or terms. In each of the cases where word frequency is plotted, the right is the long tail, to the left are the few words or terms that dominate the analysis (also known as *the 80/20 rule*).

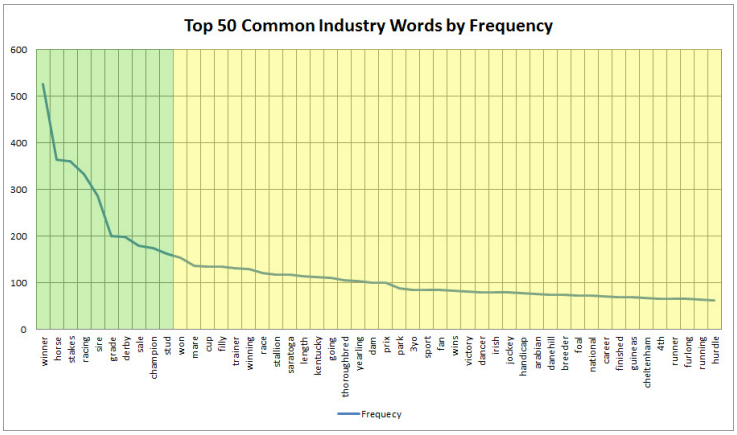


Figure 56 Top 50 Common Industry Words by Frequency – displaying features of the 80/20 rule

The Top 75 Common Industry Words, from the a combined dataset of the previously plotted Racing and Breeding words, was used to visualise the emerging Industry Folksonomy in the form or a Treemap. A treemap is an area-based visualisation which divides a rectangle into hierarchical categories, displaying relationships among large numbers of components. This provides a visual overview of the whole dataset. In the context of this research, it can be seen that the larger categories, or more frequently used words, would contribute towards the emerging industry folksonomy and highlight key concepts and classes that would form the basis for developing a basic domain specific ontology.

Figure 57 Emerging Industry Folksonomy - Top 75 words combining Racing and Breeding feeds (see Appendix C)

6.4 Ontology Development

Folksonomies and Ontologies have long been regarded as competitors and the advent of collaborative tagging has inspired lengthy global debate about which can be the best approach to categorisation. Limpens, Gandon and Buffa (2008) propose that there is a need for the users of Social Software to agree about the Knowledge Representations that support their collaborative categorisation and organisation of information and knowledge.

As previously discussed, folksonomy development is often regarded as bottom-up, while development of a formal ontology is top-down. Limpens, Gandon and Buffa (2008) argue that opposing folksonomies and ontologies in this way is counterproductive and they examine the potential for combining both approaches towards development of a community based knowledge representation. They hypothesise that the synergy of both folksonomies and ontologies can produce more intuitive knowledge representation, which includes both formal machine readable representations and informal representations which the community users can make sense of.

This forms the basis for a symbiotic relationship where both the community folksonomy and the derived ontology can work in a cyclical partnership to expand, strengthen and formalise the knowledge represented in the community. The following section describes the steps taken to develop a basic domain-specific ontology, be identifying and formally modelling key concepts from the emerging Industry Folksonomy.

6.4.1 Development Methodology

In order to begin the process of ontology development an adequate development methodology had first to be identified. Section 4.6.2, of this dissertation, outlines a proposed seven step development methodology to aid the process of a first time ontology development. It was found that the first 3 steps in the development methodology could be ignored for the following reasons:

1. The domain and scope of the ontology has already been identified within the dissertation.
2. No existing domain-specific ontologies have been found that could be reused.
3. The preceding section, 6.3.4, has identified key terms to be used in the ontology, through the emergence of a basic Industry Folksonomy.

By employing the remaining 4 steps of the methodology, the selection of a class, its properties and instances were created and represented as described in the following section.

6.4.2 Basic Domain-Specific Ontology

The rationale behind the use of the described methodology was to guide the extraction of relevant terms, from the Industry Folksonomy, which could then be used to develop a "lightweight" or basic domain-specific ontology. In the context of this dissertation it was not feasible to develop a complete ontology so a specific concept, or class, was selected which would demonstrate the process and resultant ontology.

'Horse' was selected as the concept to model and represent as a simple ontology. Using the Industry Folksonomy, identified in Section 6.3.4, a number of additional properties and classes were identified, as well as other more obvious properties required to model the concept of a Horse.

The key terms, additional properties and their relationships were identified as follows:

Class	
Thoroughbred Horse	
Relationship	**Property**
is named	Name
is	Sex
sired by	Sire
out of	Dam
trained by	Trainer
bred by	Breeder
is a	Champion

Table 9 identified Class, properties and relationships for Horse

Once identified, the class and its properties were visualised using a Knowledge Model created using CmapTools, a concept and knowledge mapping tool developed the Florida Institute for Human & Machine Cognition. This provided a hierarchical, human readable, visualisation of the Thoroughbred Horse class that would be used to begin development of the basic domain-specific ontology.

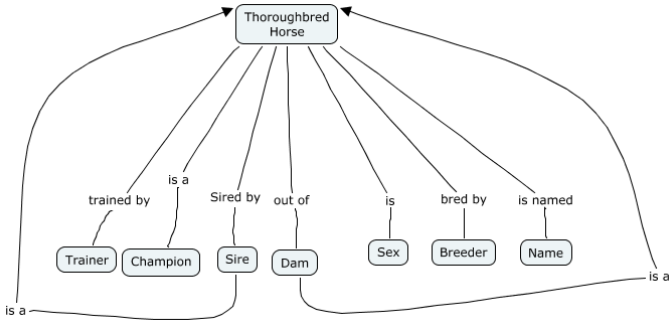


Figure 58 A simple knowledge model for a Horse, based on identified industry terms

The knowledge model presents the Thoroughbred Horse as the root concept identified as a core term used by both the Racing and Breeding domains. A number of industry terms and concepts, associated with a Thoroughbred Horse, are represented using their relationship to the root concept. These related terms and relationships have been identified earlier in Table 9.

Further, subclass relationships were identified and depicted in the knowledge model, e.g. Sire and Dam are both subclasses of the root class (they are both Thoroughbred Horses). Other properties could, similarly be interlinked as subclasses of other Industry concepts, such as Trainer and Breeder. Both are people and would have a subset of properties including, Name, Location, etc. It was decided, however, that for the purpose of this dissertation

The basic domain-specific ontology for the Thoroughbred Horse class was developed in Altova SemanticWorks, using the OWL Lite ontology development language.

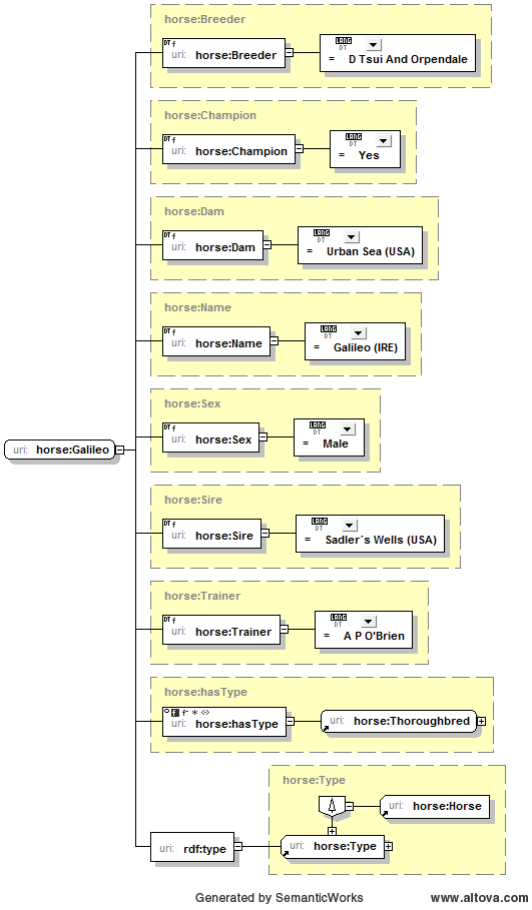


Figure 59 Basic OWL Lite construct for the Horse class of an individual Thoroughbred Horse - Galileo (IRE)

The idea behind the OWL Lite language is that it provides a minimal useful subset of ontology language features, which are relatively straightforward for ontology developers to model. The language constructs of OWL Lite provide the basics for subclass hierarchy construction including, subclasses and property restrictions. In addition, OWL Lite allows properties to be made optional or required. This minimal language subset was found to be more than adequate for developing the class, properties and instance for a Thoroughbred Horse, as required for the basic domain-specific ontology.

Figure 59, shows the expanded instance of the developed Thoroughbred Horse class, for 'Galileo (IRE)', a Champion 3 year old in Europe, now standing at Coolmore Stud. It shows that the instance class for Galileo is of type 'Thoroughbred' and that it is a subclass of the 'Horse' class. The properties of the Thoroughbred Horse class are instantiated for Galileo and the appropriate information recorded in each, based on

information available from the racingpost.com and Coolmore stud websites, see Figure 60.

Figure 60 Details of Galileo (IRE) as found on the Coolmore Stud website

Figure 61, provides an RDF representation of the instantiated class for the Thoroughbred Horse, 'Galileo (IRE)'. It displays an XML representation of the class properties, its domain class (the root class from which it is instantiated) and the values for each of the properties.

Figure 61 Example of RDF representation of a Thoroughbred Horse Instance (OWL Lite) based on terms found in the Industry Folksonomy

Although, in the context of this dissertation, the ontology represents an instantiation of a Thoroughbred Horse, the ontology has been developed as a domain independent representation. The throughbred type is also an instantiation of the superclass of Horse, upon which the example is based. Similar instantiations could be created for other breeds of Horses. The complete OWL Lite development is included on the accompanying CD for further examination.

6.4.3 The Semantics

Using a centralised basic domain-specific ontology, the collaborative tagging interface of a community web portal could be strengthened and improved. The idea is to disambiguate while tagging, by suggesting to users relevant connected terms guided by the newly developed ontology.

When a concept does not exist, users are free to propose a new one to the administrators of the portal, who in turn will include it in the correct context in the ontology. Social tagging and folksonomy development can be seen in this instance as an empowerment of the development and upkeep of the domain-specific ontology, which in turn helps formalise and disambiguate the tags and folksonomy.

This process can be further strengthened by using Semantic Web frameworks, such as SIOC (Semantically Interlinked On-line Communities) and FOAF (Friend of a Friend) to develop a Semantic Bootstrapping process for the Industry-based community tagging.

Semantically Interlinked On-line Communities (SIOC)

The Semantically Interlinked On-line Communities (SIOC) project, developed by DERI, provides developers of social Web platforms with a formal and technological ontology framework to describe the resources exchanged within and across the online community.

SIOC describes the most common elements present on Web Sites of communities: the concept of ‘site’, the concept of ‘post’ of a weblog, the concept of ‘user’, etc. According to Bojars, Passant, Breslin and Decker (2010) SIOC aim is to provide a

comprehensive data model based on Semantic Web technologies in order to represent online communities and their activities in a homogenous way.

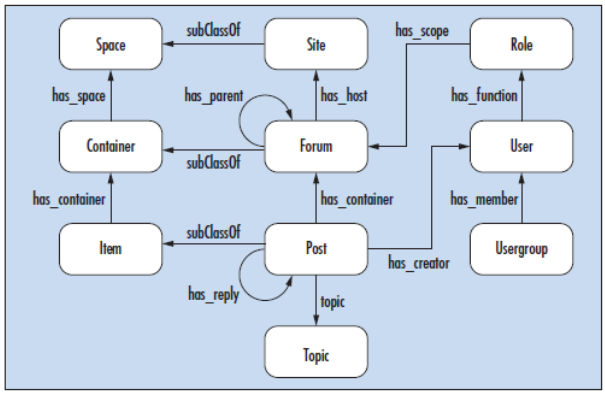


Figure 62 The main classes and properties in the Semantically Interlinked Online Communities (SIOC) ontology.

The SIOC ontology class for User, sioc:User, is related to the 'Agent' class from the Friend of a Friend (FOAF) ontology and as such reuses the FOAF vocabulary to describe person-centric data and relationships (Bojars et al., 2008).

Friend of a Friend (FOAF)

Friend of a Friend (FOAF) provides an RDF/XML vocabulary to describe personal information, including name, mailbox, homepage URL, friends, and so on. The FOAF vocabulary defines both classes (e.g., foaf:Agent, foaf:Person, and foaf:Document) and properties (e.g., foaf:name, foaf:knows, foaf:interests, and foaf:mbox) grounded in RDF semantics.

FOAF documents in turn stimulate a "web of acquaintances", or social network, an implicit trust network to support such applications as knowledge sharing within online communities. Using FOAF, machines can recognise a users home pages, weblogs, wikis, etc. This in turn forms the basis of the social network as the relationships that connect people, places and things described on the Web, are exposed to be read and interpreted by machines and online social applications.

```
<rdf:RDF>
  <foaf:Person>
     <foaf:title>Mr</foaf:title>
     <foaf:nick>Mick</foaf:nick>
     <foaf:surname>Hartnett</foaf:surname>
     <foaf:firstName>Michael</foaf:firstName>
     <foaf:mbox>michael.hartnett2@student.dit.ie</foaf:mbox>
     <foaf:schoolHomepage rdf:about="www.comp.dit.ie"/>
    <foaf:knows>
      <foaf:Person>
         <foaf:surname>Bradley</foaf:surname>
         <foaf:firstName>Ronan</foaf:firstName>
         <foaf:mbox>Ronan.Bradley@dit.ie</foaf:mbox>
       </foaf:Person>
     </foaf:knows>
    <foaf:knows>
      <foaf:Person>
         <foaf:surname>Gordon</foaf:surname>
         <foaf:firstName>Damian</foaf:firstName>
         <foaf:mbox>Damian.Gordon@dit.ie</foaf:mbox>
       </foaf:Person>
     </foaf:knows>
   </foaf:Person>
 </rdf:RDF>
```

Figure 63 Simple FOAF RDF file

Tying the Bootstraps

The proposed Bootstrapping process will be based on existing research carried out in the field of the Semantic Web. Torben Knerr (2007) proposes a Tagging Ontology which uses the FOAF vocabulary to represent the user profiles of community taggers, allowing each user to keep their tagging data in one place. This forms the basis for a community based tagging folksonomy.

Passant (2007), of the Digital Enterprise Research Institute (DERI) at the National University of Ireland, Galway, discusses the benefits of using the SIOC ontology to "Strengthen Folksonomies and Enrich Information Retrieval in Weblogs". Passant proposes the application of the SIOC ontology to help with tag disambiguation, by using the sioc:topic relationship to contextual the tags. Also by leveraging the SIOC associations with other ontologies tracking of the original tags for blog posts can be maintained, using FOAF to indicate the originating tagger.

Passant (2007) further discuss various applications of Semantic Web technologies to online social spaces allowing for the expression of different types of relationships between people, objects and concepts. By using common, machine-readable ways of expressing individuals, profiles, social connections, and content, they provide a way to interconnect people and objects on the Web in an interoperable, extensible way.

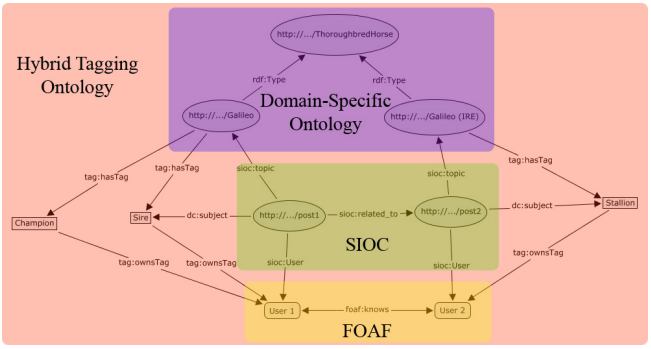


Figure 64 Hybrid Tagging Ontology, used to bootstrap the Folksonomy and Ontology integration

Figure 64, depicts an envisaged hybrid tagging ontology based on the earlier developed domain-specific ontology, and incorporating the SIOC and FOAF ontologies. Each of the individual ontologies and their relationships can be clearly seen, with each contributing as an individual element to the hybrid ontology. The use of the basic domain-specific ontology would be to instantiate classes for the key terms and concepts in the Industry Folksonomy, forming the start of the Semantic formalisation process.

The SIOC ontology will formalise the relationships between the various posts, by various users, that refer to topics inferred from the domain concepts. This level of the hybrid system would serve to disambiguate the concepts by establishing the sioc:related_to property between the posts. In turn the SIOC ontology can be seen to interact with the FOAF ontology to determine relationships between users. These relationships form the basis of the social network within the community, establishing an infrastructure of trust and discovery, allowing the system to assist user tagging by suggesting viable tags previously used by their domain associates.

The proposal of new tags, or folksonomy concepts, and their subsequent discovery and use by other users in accordance with the proposed hybrid system, can be seen as a bootstrapping, or self-sustaining, process between the Folksonomy and Ontology integration. The potential of such a hybrid system which exploits the benefit of both the ease of use of folksonomies and the support of the formalisms and the methods of the Semantic Web, provides new and improved perspectives for assisting Knowledge Sharing on the “Socially Semantic Web”.

6.5 Conclusions and Future Work

In this Chapter, a review of Industry RSS feeds was undertaken in order to assess their suitability for the conveyance of organised domain information and knowledge. It was found that the lack of adequate and exploitable tagging and categorisation within the syndicated feeds, meant that an alternative method of analysis would be required to develop an industry folksonomy.

Through examination of the content of syndication feeds within the community, a bottom-up process of key term identification, based on frequency of use, was conducted. The emerging terms were analysed, finding distinct correlations between the specific domains in the industry. These correlations were to be expected given the interconnected nature of both the Racing and Breeding domains in the Thoroughbred Industry and the emergent Industry Folksonomy was testament to this.

The development of a basic domain-specific ontology, based on an earlier identified methodology, was performed using a single industry concept of 'Horse'. Since the industry revolved around horses, this was viewed as the prime starting point. Additional terms and properties were identified within the folksonomy and their direct relationships with the 'Horse' class were modelled and used to produce a simple ontology. A discussion about extending this ontology using other formal, or Semantic, Ontologies formed the basis of a proposed Hybrid Tagging System that could be used to bootstrap the interaction between the industry folksonomy and domain-specific ontology.

One aspect of future works identified would consist of additional data cleansing processes. The processes described in Section 6.3.4, of this chapter could be extended to include additional domain-specific stop words, although due consideration would need to be given to the effects that their removal would have on the emergent industry folksonomy. Since the Thoroughbred Industry tends to use a more descriptive terminology, concentrating more on the familiarity of the community with names of Horses, Trainers, Breeders, Studs, Racecourses, etc., removal of these additional terms could destabilise the folksonomy. The additional domain-specific stop words might include:

- Country Names
- Racecourse Names
- Horse Names
- Names of People including; Owners, Trainers, Jockeys, Breeders, etc.
- Names of Organisations, Studs, etc.

A second area for future work would involve the analysis of term variance over time. This form of analysis was deemed not applicable in the context of this dissertation as the experiment was conducted in order to acquire a current snapshot of the key terminology within the industry to create a bootstrapping process. The terminology can be viewed as fairly static as the industry has been established for more than 250 years, but further examination would be used to support this and also identify if cross-culturalisation and globalisation of the online industry have had an effect on the industry language.

7 DESIGN AND EVALUATION

7.1 Introduction

The focus of this chapter will be used to present, discuss and evaluate a prototype for a potential Thoroughbred Industry Knowledge Portal. A dual purpose prototyping methodology will be employed to firstly design a low fidelity prototype ("Paper Prototyping") and secondly evaluate the prototype ("Wizard of Oz Prototyping") allowing the selected evaluation audience to observe the functionality of the prototype without a requirement to be proficient in its use.

The design and usability evaluation will be conducted as a successor to the findings and proposals in the previous chapter, with two main concepts being modelled; Social Networking and Communal Tagging. Feedback from the evaluation audience will be presented and analysed to provide conclusions of the relevancy of design, the ease of use and functionality and the potential for further development.

7.2 Design and Evaluation Method

The design and evaluation process follows the Paper Prototyping method, which is often used as the first step of rapid prototyping. Snyder (2001) describes Paper Prototyping as a method of design and "*usability testing where representative users perform realistic tasks by interacting with a paper version of the interface*". This method of has been championed by several leading usability experts including Jakob Nielsen, indicating that "*any mid-sized design project will probably get an ROI of several thousand percent*" through the use of Paper Prototyping.

Paper Prototyping is particularly useful for confirming the legibility of an initial design for a Web Site or Portal using a low-fidelity design mock-up of the appropriate pages. Each page is presented to a user to test the perceived functionality, legibility and usability. Paper prototyping is also the recommended design testing technique in the contextual design process.

"*The Contextual Design methodology, developed by Karen Holtzblatt and Hugh Beyer, is a customer-centred design process which uses extensive field data as the foundation for understanding users' needs, tasks, intents, and processes in order to design products and systems that meet both users' and business' needs.*" (InContext, 2011) In practice, this means that researchers aggregate data from customers in the field and apply these findings into a final product. This information is captured by observing the users behaviour and conversing with the user while he or she evaluates the prototype. A key aspect of the technique is to connect with the user, letting their work and the issues they encounter guide the evaluation.

There are three main techniques of Paper Prototyping, used for design and usability testing:

- *Compositions* – Visual representations of websites used to demonstrate various aspects of the interface.
- *Wireframes* – Visual representation of websites used to demonstrate the page layout of the interface.
- *Storyboards* – Visual representation as a series of images or interactive pages used to demonstrate the functionality of the interface.

For the purpose of this research a combination of wireframing and storyboarding will be used to provide the test user with a visual page layout with certain interactive functionality, allowing the user to click through from page to page.

When Paper Prototyping is used for evaluation it is sometimes referred to as "Wizard of Oz Prototyping" since an instructor ("Wizard") much of the functionality of the system and makes choices in terms of the next page presented to the user that the working system when completed would do. (Bernsen, Dybkjær and Dybkjær, 1993)

The following methodology will be employed throughout the remainder of this chapter in order to design and evaluate the usability of the prototyped Thoroughbred Industry Knowledge Sharing Portal, and finally aggregate and present the user testing findings:

1. Design Paper Prototype concentrating on scoped features including Signup and Friend Discovery (FOAF) and Tagging.
2. Use Wizard of Oz Prototyping to evaluate usability of prototype with target audience.
3. Present findings of usability testing and propose areas for improvement.

7.2.1 Scope

Paper Prototyping is commonly used for representing simple concepts and as such are ineffective for demonstrating complete software application or web sites. For this reason it is necessary to limit the scope of the prototype, to be designed and evaluated, by focusing on the identifiable usability issues and requirements. During the early stages of the conceptual design phase, it is more practical to concentrate on the general design of a system, rather than the specifics. This is a useful indicator of how well the prototype design meets the user's requirements and expectations.

For the purpose of this research, the scope of the prototype design and evaluation focused on the two main concepts described in section 6.4.3 of the previous chapter:

- Social Networking through the concept of Friend of a Friend (FOAF), and
- The concept of Communal Tagging within the portal using Semantically-Interlinked Online Communities (SIOC) to build on the FOAF relationships.

Each of these concepts will be modelled in the conceptual design of the prototype and evaluated using the selected user audience.

7.3 Prototype of HorseE

A conceptual Knowledge Portal prototype, called HorseE, Figure 65, was designed using Balsamiq Mockups, for user testing and proof-of-concept of the proposed Bootstrapping techniques described in the previous chapter (see Section 6.4.3). The basic concept employed when designing the prototype for HorseE was a simple, clear layout used to disguise powerful Knowledge Management and Knowledge Sharing functionality. The design principle is to conceptualise a deliverable portal that can

accommodate users of all industry status, i.e. expert, professional, amateur, spectator, etc., and all levels of technical competency, i.e. low to high computer and web literacy.

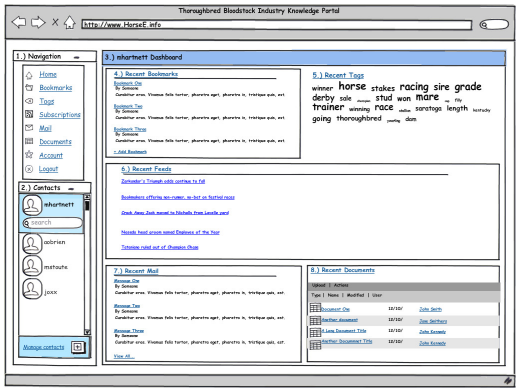


Figure 65 "HorseE" Dashboard - Conceptual Prototype Design for Thoroughbred Industry Knowledge Portal

The conceived portal design combines elements inspired by numerous existing, and popular, web based portals.

1. *Navigation* – A simple and clear navigation menu, similar to Google.
2. *Contacts* – A manageable Contacts/Friend List, analogous to Google Contacts or Facebook Friends.
3. *Dashboard* – A central account dashboard for displaying informational dashlets, etc., common to all web based portals.
4. *Recent Bookmarks* – A shareable repository of favourite bookmarked websites, blogs, etc., akin to Delicious.com.
5. *Recent Tags* – This Tag cloud represents the most recent and frequently tags applied by the HorseE user.
6. *Recent Feeds* – A shareable repository of favourite blog articles and feeds similar to Digg.com.
7. *Recent Mail* – A mail facility, providing the HorseE user with the ability to send and receive mail from multiple registered mail clients, e.g. GoogleMail, Hotmail, etc.
8. *Recent Documents* – A repository of documents, allowing the user to store and share documents with other connected friends/users on the HorseE portal, comparable to Sharepoint or Google Docs.

Each of the conceptualised elements was chosen based on two main premises:

1. Their applicability and functionality within in the Knowledge Portal. Each of the elements can be leveraged to provide a constant source of information and knowledge that is consumable and shareable by the user.
2. Familiarity of the expected users with the elements was a critical factor. It was perceived that familiarity, and confidence with the use of these elements would provide an easy and continuous user experience, based on current web trends.

7.3.1 Social Networking

The envisaged User Signup process for the HorseE Knowledge Portal will mirror many existing portal signup processes, e.g. Delicious, Yahoo's Social Bookmarking Portal. This common model for portal signup and sign in provides users with various methods for accessing information and knowledge contained within the portal. Users can create a new user account, including User ID and Password, specific to the portal or by signing in using an existing Social Networking profile.

Social Networking provides the basis for the knowledge sharing processes employed within the HorseE Knowledge Portal. By providing users with the ability to sign in using Social Networking profiles, they can import their existing contact lists in order to search and discover other HorseE users with whom they might have a social relationship. Additionally, the use of existing Social Networking profiles will also allow the user to import their existing bookmarks, RSS feeds, email and documents, shared via their profile, for use within the HorseE Knowledge Portal.

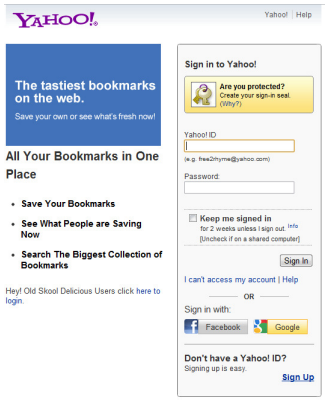


Figure 66 Signup / Sign in screen for Delicious, Yahoo's Social Bookmarking Portal

As can be seen a number of similarities exist between the signup/sign in page for Delicious, Figure 66, and the signup/sign in page for HorseE, Figure 67. Both portal pages allow users to create and sign in using a specific HorseE user profile and also

provide a sign in facility using existing Social Networking profiles. One significant difference is the inclusion of a HorseE Knowledge Cloud; see Figure 67, which updates on a daily basis.

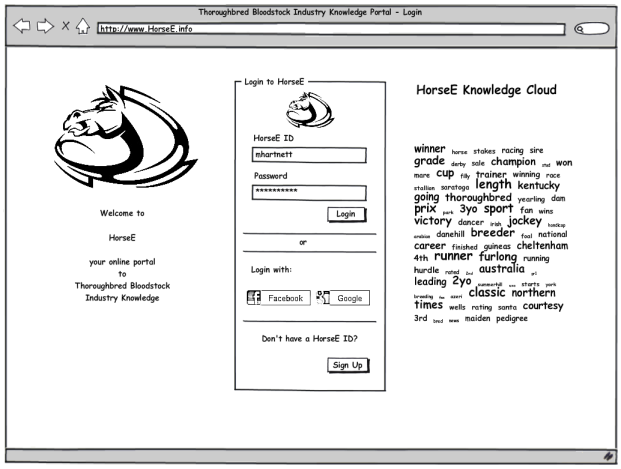


Figure 67 Signup / Sign in screen for HorseE Knowledge Portal

The HorseE Knowledge Cloud is provided as an initial method of Knowledge Sharing within the HorseE Knowledge Portal, functioning as a dual purpose tool. Its primary function is to provide a visual representation of the most common tags used by HorseE Knowledge Portal users to categorise the Knowledge within the portal. The duality of the Knowledge Cloud becomes apparent as it can firstly be used by existing users to find, bookmark, share and tag, articles tagged using words represented in the Knowledge Cloud. Secondly it can used by non-HorseE users as a means of finding and reading articles tagged using industry concepts by industry members. In this second scenario, the reader does not require a HorseE user profile, or be signed in, but the ability to bookmark, share or tag the articles is not available.

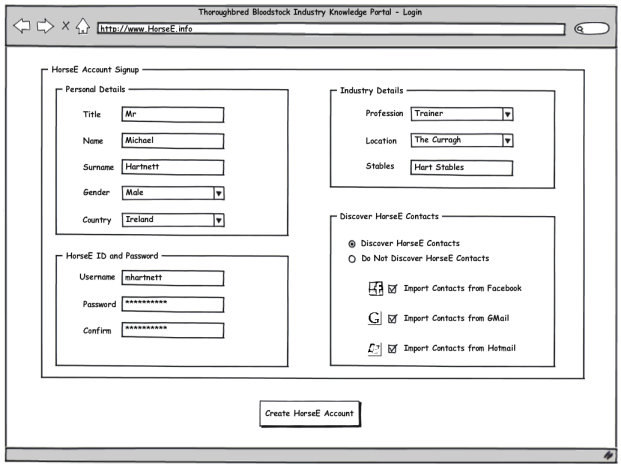


Figure 68 HorseE User Sign Up Page

The HorseE user signup process is used to create a new HorseE profile, which can be used for the purpose of Knowledge Discovery and Sharing within the portal. The Account Signup page is used to gather information about the new user for a number of purposes:

- To provide the user with a unique HorseE Username and Password.
- To determine the area of the Industry, the user is involved in.
- To create a Friend-Of-A-Friend (FOAF) profile and discover existing FOAF contacts.

The creation of a FOAF profile provides the impetus for the generation and expansion of a semantically interlinked network of users, see Section 6.4.3, forming the basis for a community of taggers who will in turn produce a community based tagging folksonomy for the industry. This semantically interlinked network of users is created by allowing HorseE users to search for and save other HorseE users as contacts within their profile.

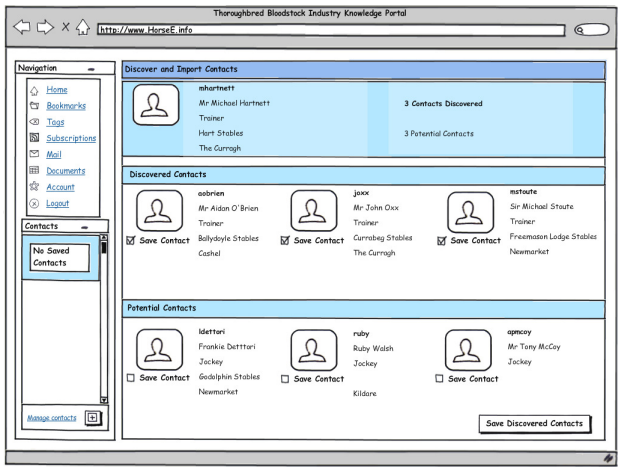


Figure 69 Discovery of existing HorseE users, with FOAF profiles, using imported Social Network contacts

To bootstrap the process of contact discovery, the new user can import their contacts from existing Social Networking profiles, i.e. Facebook, Google, etc., in order to discover if any of them are HorseE users. Since each of the discovered contacts will also have an existing HorseE FOAF profile, these can be linked using the 'foaf:knows' property, indicating a reciprocated interaction between the HorseE users.

Finally, the creation of a semantically interlinked network of HorseE users, allows each user to discover the RSS Feeds subscribed to by their counterparts. This discovery process can be used at initial signup in order to identify the areas of information and knowledge which other industry members are interested in. As well as importing subscriptions from their existing Social Networking profiles the HorseE users can, at any time, import the subscriptions of their HorseE contacts, see Figure 70. This facilitates one of the Knowledge Sharing initiatives conceived as part of the HorseE Knowledge Portal.

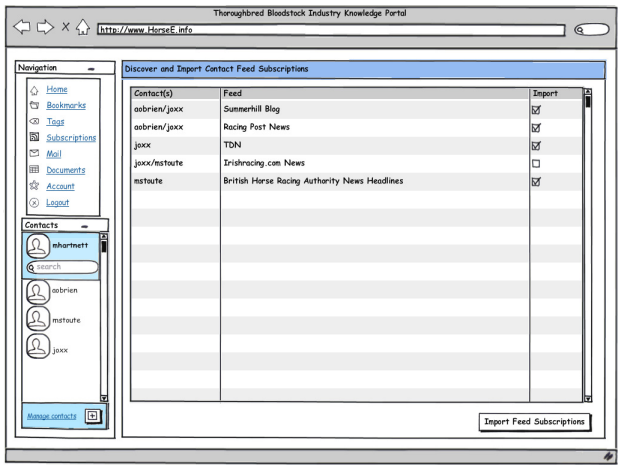


Figure 70 Discovery of feed subscriptions based on contact subscriptions

7.3.2 Communal Tagging

The second bootstrapping concept to be modelled using the HorseE Knowledge Portal prototype is that of Communal Tagging and the relationships between the HorseE users and the community tags. This requires the introduction of the SIOC ontology, coupled with the extant FOAF profiles, to form the basis of the social semantic community which can be leveraged for the purpose of Communal Tagging.

As previously discussed, in Section 6.4.3, the SIOC ontology will formalise the relationships between articles and users and will establish a system of Communal Tagging within HorseE, through an infrastructure of trust, discovery and suggesting viable tags previously used by portal users.

Figure 71 shows the paged feed subscriptions for a HorseE user. Each article presented as part of the feed subscription, displays as a HTML representation of the article with inline images, links and rich text similar to any other RSS Reader. The HorseE user is provided with a number of features allowing them to bookmark the article, unsubscribe from the article's feed, comment on the article, and most importantly in the context of the HorseE Knowledge Portal, the ability to Tag and Share the article.

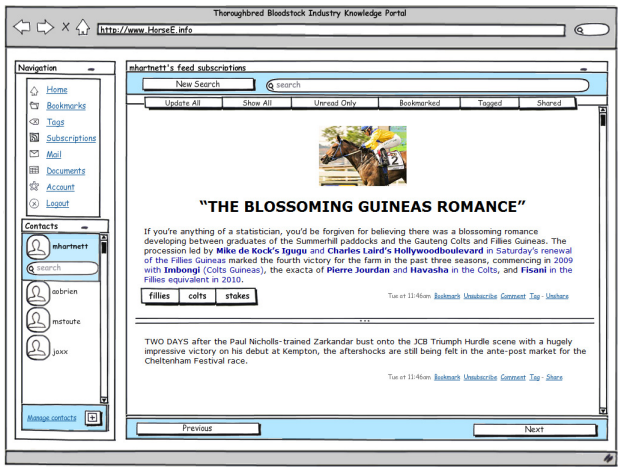


Figure 71 Subscribed Feed Article with Shared Tags

The tags associated with each article are prominently displayed below the article. The concept of Communal Tagging is introduced when a HorseE user adds their own Tags to an article and then Shares that article with their contacts.

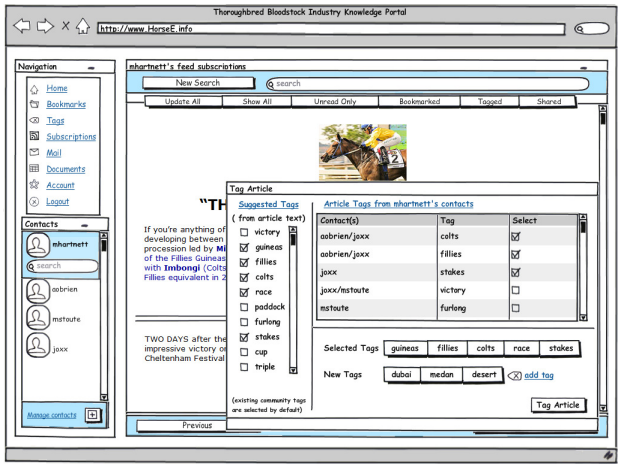


Figure 72 HorseE's Communal Tagging implementation

When tagging an article the HorseE Knowledge Portal attempts to assist the user by employing a method of suggestive tagging based on identified industry concepts

within the article (determined by the domain specific ontology, introduced in Section 6.4.2, and the Industry Folksonomy in Figure 57). Additionally, the application of the SIOC ontology in this instance serves as an aid in the disambiguation of the identified concepts and tags, by discovering and displaying the tags used by the HorseE users contacts, indicating which are common to multiple contacts.

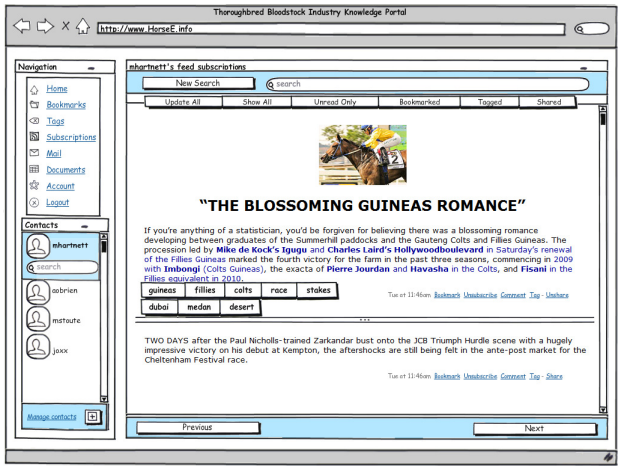


Figure 73 Newly Tagged Feed Article

The user can opt to choose from the suggested tags or add their own tags to the article. By selecting an extant tag the user is emphasising the importance of that Tag in relation to the article. The HorseE tagging engine will record this and display the weighted tag accordingly in its Knowledge Cloud. Alternatively, if the user chooses to add a new tag to the article, this tag will be ingested by the tagging engine and exposed as a shared tag, with the article, through the Communal Tagging process.

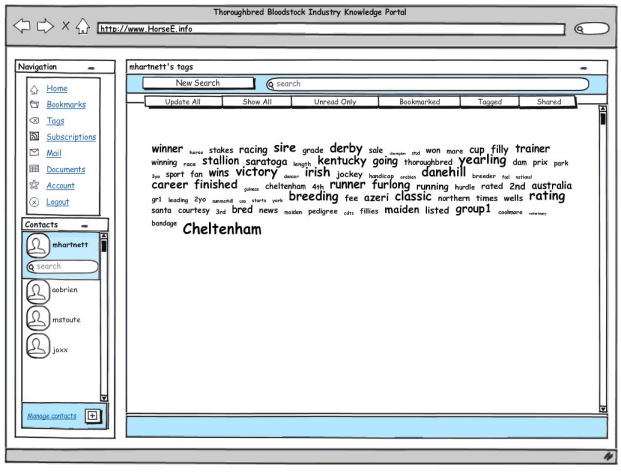


Figure 74 HorseE users Personal Knowledge Cloud

At any point the HorseE user can navigate to a portal page which will display a Personal Knowledge Cloud, see Figure 74, relating to their currently subscribed feed articles. Unlike the Knowledge Cloud displayed on the HorseE Knowledge Portals sign in page, the Personal Knowledge Cloud is used to display a weighted cloud of tags for the users subscribed feed articles.

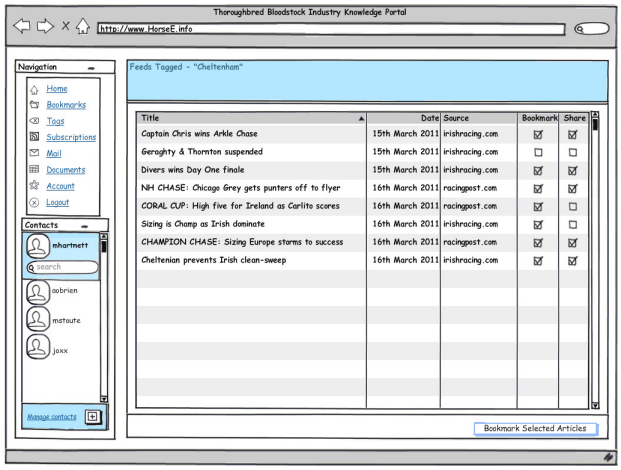


Figure 75 HorseE users Feed Articles tagged with "Cheltenham"

By selecting and clicking on a tag in their Personal Knowledge Cloud the HorseE user will navigate to a new portal page displaying a list of all the subscribed feed articles that have been tagged with the selected word. The HorseE use can then select the Title of the feed article to display and read the article. Alternatively, each of the articles can be bookmarked and/or shared to expose the user's tags relating to the shared article. This helps to continue the cycle instigated by the portals Communal Tagging system, which provides the stimulus for the self-sustenance of the tagging system.

7.4 Evaluation

The HorseE Knowledge Portal Paper Prototype was designed as a proof-of-concept for the proposed semantic solutions in Section 6.4.3. The paper prototype was reviewed with the target audience, see Section 7.4.1, and their reactions, comments and suggestions were recorded.

The comments and suggestions are compiled in the following sections providing an evaluation of the acceptability and usability of the portal concept, the conceived functionality and the look and feel of the portal.

The evaluation followed the "Wizard of Oz Prototyping" method where the functionality was simulated, and the modelled concepts explained by the researcher, i.e. the "Wizard". The interviewees in turn observed the simulation and provided an evaluation through means of recorded feedback.

7.4.1 Evaluation Audience

For the purpose of this dissertation, three participants were selected and invited to evaluate the portal prototype and provide feedback via semi-structured interviews. The semi-structured interview technique concentrated on the two main concepts being modelled in the prototype, allowing each interviewee to answer and provide observations based on their levels of understanding of the concepts.

The audience considered for the evaluation of the HorseE Knowledge Portal Paper Prototype were the following:

	Gender	Code in dissertation
Interview 1	Female	F1
Interview 2	Male	M1
Interview 3	Male	M2

Table 10 Evaluation Audience

F1 is a female subject, mid-thirties, technically proficient with the use of PC, Web, Blog and Portal technologies, with a significant insight into the thoroughbred bloodstock and insurance industries. M1 is a male subject, mid-fifties, with competent PC and Web skills, but limited proficiency with the use of Blogs and Portals, with a significant insight into the thoroughbred bloodstock and banking industries. M2 is a male subject, late-thirties, conversant in the concepts of Knowledge Management, Knowledge Sharing and other concepts being modelled in the prototype. M2 also has advanced proficiency with the use of PC, Web, Blog and Portal technologies, with a nominal insight into the thoroughbred industry.

7.4.2 Initial Impressions

In this section a list of terms and phrases that have been elicited from the interviewees, in an attempt to evaluate their initial impressions of the HorseE Knowledge Portal prototype are presented in Table 11.

Interviewee(s)	Term / Phrase	Relevant to
F1, M1, M2	Clear simple design	Design and Layout of Portal
F1	Clean	Uncluttered screens
M1	Simple	Account Signup
F1, M2	Clear	Dashboard Layout
F1, M2	Social Network	Signup, Contacts, Tagging
F1, M2	Familiar	Concept of Social Networking
F1, M1, M2	Logical	Processes and functionality within HorseE
M2	Positive	ability to use multiple sign in methods
M2	21st Century	ability to use multiple sign in methods

Table 11 Terms and phrases elicited from interviewees indicating first impression of HorseE Knowledge Portal prototype

These initial reactions and responses from the interviewees can be seen to be positive and encouraging, indicating that the HorseE Knowledge Portal prototype has attained an initial acceptance with its potential audience and user base.

The following sections will loosely apply the PMI technique developed by Edward de Bono to categorise the comments and reactions of the interviewees in order to outline the perceived "Pros and Cons" (Positives and Negatives), and interesting observations (Interesting) that have emerged through evaluation of the HorseE Knowledge Portal prototype.

7.4.3 Positive Evaluation

This section categorises the "Positive" interviewee experiences gathered from the evaluation of the HorseE Knowledge Portal prototype. These "Positive" experiences were based on the perceived success of implementation of the specified concepts and their functionality within the prototype.

Interviewee	Positive Comments
F1	Easily recognisable Social Networking concept, familiarity is extremely important factor in decision to use the portal.
F1	Easy to understand tagging process, concepts used for communal tagging and disambiguation were good.
M1	Bookmarking of feed articles is important for identifying and registering potential business.
M1	Tags provide ability to reach and capture a wider group of people with different viewpoints on the same topic.
M1	Mail facility effective method for reaching a wider group of potential business contacts.
M2	Good clear interface, with the initial tag cloud adding push/pull benefit by providing something for free, for users and non-users.
M2	Strong Social Networking concepts – well executed.
M2	Functionality is not restricted if all personal/profession details are not filled in.
M2	Registration process does not post to other Social Networking sites, from which contacts are being imported. Respects privacy.
M2	Good idea for envisaged method of subscription from blogs, e.g. the user of a HorseE icon, similar to  .
M2	Mail Service increases the range of people using portal and allows for those users not adept with using Social Networking facilities/concepts reach a wider range of users. (Mirrors comment from M1).

Table 12 Positive Evaluation Comments from Interviewees

It can be inferred from these positive responses that the interviewees felt comfortable with the modelled concepts and the methods through which their functionality was represented in the prototype.

7.4.4 Identified Issues

This section categorises the issues identified by the interviewees experienced during the evaluation of the HorseE Knowledge Portal prototype. These issues do not necessarily correlate with "Negative" experiences, but rather serve as functionality that is lacking in the current inception of the HorseE Knowledge Portal prototype.

Interviewee	Identified Issues
F1	Need to be able to invite new users. This will increase user base of the portal, which in turn will increase the knowledgebase.

F1	Need to be able to share feeds by invitation as well as having a discovery process for the feeds.
M1	Need some mechanism to restrict subscriptions which are specific to the Thoroughbred industry, e.g. a self-updating repository of feeds identified by web crawling agents.
M1	The Thoroughbred Industry is a Global Industry – How will multilingual feeds and users work?
M2	Tag Clouds will require upper limit, i.e. approx. 5 occurrences, in order to avoid any one tag overwhelming the others in the cloud.
M2	There is a perceived lack of privacy settings within the portal, to restrict visibility to other portal users.
M2	When a user first registers, the dashlets should not be blank, but provide information for users so they are not confused when information suddenly appears.

Table 13 Comments from Interviewees identifying issues with the prototype evaluation

The identified issues provide insight into the potential deficiencies within the HorseE Knowledge Portal prototype. These issues also offer an indication of some of the concerns that the interviewees, and potential users, have in respect of privacy and possible confusion arising from lack of informative messages.

7.4.5 Interesting Comments, Suggestions and Observations

This section seeks to present the "Interesting" aspect of the PMI technique, by identifying comments, suggestions and recommendations, which the interviewees offered in respect of improvements and extensions for the HorseE Knowledge Portal prototype.

Interviewee	Interesting Comments, Suggestions and Observations
F1	Ability to add new contacts based on the tags they have suggested on an article.
F1	Forum – This could provide the impetus for a conversational method of Knowledge Sharing. Also potential for an industry Golden Pages.
M1	Golden Pages for Thoroughbred Industry. (Mirrors comment from F1)
M1	Market Place that can be used for promoting business.
M2	Additional flexibility could be provided by greater user configurability, e.g. ability to show or hide tags in personal Tag Cloud.
M2	Discovery process for feeds, suggested based on article tags that you did not use.
M2	Ability to import picture/avatar and personal details from other social network profile – to keep all profiles uniform.
M2	Ability to be aware that other users are currently logged in – Instant Message chat, i.e.

	Gmail.
M2	Quick way of being able to add blog subscription via special qualifier code tags.
M2	Provision of a simple method for agreeing with, or confirming, Tags on a feed article without having to add another tag.
M2	Rewards System for Tagging.

Table 14 Interesting Comments, Suggestions and Observations provided by the interviewees

It can be seen from the comments in Table 14 that interviewee M2 has a greater understanding of the concepts of Knowledge Management and Knowledge Sharing by virtue of the comments and suggestions provided. It is also evident that while both F1 and M1 have significant insight into the Thoroughbred Bloodstock Industry, the suggestions provided by both regarding an Industry Golden Pages display a distinct understanding of the business and marketing potential displayed by the HorseE Knowledge Portal prototype.

7.4.6 Analysis of feedback

As previously inferred in Sections 7.4.2 and 7.4.3, the positive responses provided by the interviewees indicate that the HorseE Knowledge Portal prototype has been accepted as a potential Knowledge Sharing technology by its potential audience and user base. Inference can also be made that Social Networking and Communal Tagging concepts have been successfully modelled in the prototype, based on the positive feedback from the interviewees with regard to the functionality and ease of use of each concept.

Two thirds of the interviewees suggested that the discovery methods used for identifying both contacts and feeds lacked depth and were limited in its functionality. Both interviewees indicated the potential for extension of the discovery methods to identify relevant feeds based on tags on existing subscription articles and also for potential expansion of the users contacts based on those tags. In the context of the research these improvements would serve as an extension of the Social Networking relationships within the portal, which in turn would lead to an increase in the Knowledge Sharing facilities within the portal community.

One of the interviewees indicated the potential for one single tag having the ability to overwhelm the other tags contained within one of the numerous tag clouds displayed within the portal. A suggestion that an upper limit of tag occurrences be applied, to the weighting of the tags within the cloud, in an attempt to counteract overpowering effect. In the authors view, this upper limit, while practicable, could potentially cause a stagnation of the tag clouds where all tags occur more than the specified upper limit. In this case all tags would be displayed with the same weighting. A better solution may be to apply an upper limit of occurrence in conjunction with a coloured or shaded palette allowing tags that exceed the upper limit of occurrence to have progressively darker shades of colour indicating their greater popularity.

An issue of significant concern with one of the interviewees was privacy within the HorseE Knowledge Portal. This concern was highlighted in the lack of perceived privacy settings allowing the user to indicate which other HorseE Knowledge Portal users can discover that the user has registered, is currently a user or has been active within the portal. The provision of a set of privacy settings, similar to Facebook for example, could be implemented allowing the user to indicate their level of visibility with the additional facility of specifying an exclusion list of users they wish to remain invisible to.

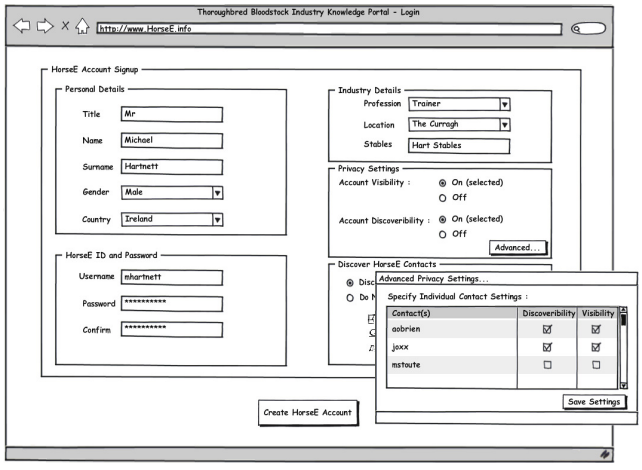


Figure 76 Revised Signup / Account Settings with Privacy Settings

Another issue highlighted by the prototype evaluation is the potential confusion for users, when they first register with the portal. If no contacts, feed subscriptions and ultimately no tags are saved in the user's profile, then the users portal dashboard will contain numerous empty dashlets.

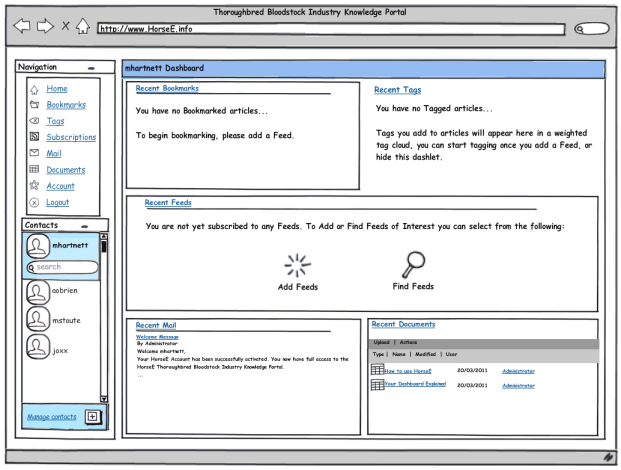


Figure 77 Revised Portal Dashboard with Informative Messages

When the user begins to add subscriptions to feeds, bookmarks and contacts, it was perceived by the interviewee that the logical assumption would be that the corresponding dashlets would populate, however all users may not perceive this functionality. The population of the Tag Cloud dashlet in particular may cause confusion unless an adequate message is displayed indicating to the user the reason why the dashlet is empty and how it will respond when they begin to tag articles. In the authors view any messages that are displayed in the empty dashlets should be an invitation to use the dashlets' functionality as opposed to an indication that the user has not provided information to populate the dashlet.

Additionally, a number of interesting suggestions were observed during the interview process. Two thirds of the interviewees saw the potential for the implementation of an Industry Golden Pages within the HorseE Knowledge Portal, providing the community with marketing and business initiatives and also the discovery of prospective relationships. This concept could be further extended to develop and maintain a "Knowledge Yellow Pages" for Thoroughbred Industry providing portal users with a concise directory and access to the knowledgeable experts within specific areas of the industry.

It was suggested by one of the interviewees that future research and development of the HorseE Knowledge Portal may provide a quicker mechanism for subscribing to feeds from within the portal and also for indicating agreement with existing tags on an

article without having to use the existing tagging mechanism. In relation to the quick subscription to feeds, the interviewee suggested that a predefined qualifier could be entered that would initiate the subscription process to a specified feed, e.g. "%Subscribe%blog.summerhill.co.za%Subscribe%". This would indicate to the portal that the user wished to subscribe to the blog feed within the qualifier. In the authors experience, this method for quick feed subscription would work well in portal implementations in IT or Computer related industry, but would not provide much in the way of improvement in a non-technical industry such as the Thoroughbred Industry.

The suggestion of a quick tag agreement mechanism would complement the existing tagging mechanism, by providing a streamlined process by which users can indicate agreement with existing article tags without having to reselect those tags again. This suggestion provides the impetus for a two tier tagging mechanism. One tier exists where the correct way for tagging currently implemented could be used by novice portal users; to aid their tagging through a folksonomy backed suggestive mechanism. The second tier will provide a more practicable mechanism, used primarily by experts, for agreeing with an article's tags and reinforcing their importance through a single click, or "Agree" icon, similar to the "Like" facility on Facebook.

Finally, another area for future consideration and research was indicated, where a reward system could be implemented for the active employment of Tagging by a user. This rewards system could take the form of a credit accumulation system, e.g. Pigsback.com, where users are rewarded with credits for simply viewing adverts or taking up offers on the website. These credits can then be used to purchase items or vouchers. In a similar fashion the HorseE Knowledge Portal could provide a credit based system, which would reward users based on their Tagging activity. In the context of the industry, these credits could be used to purchase subscription only Pedigree documents, racing videos or even Horse images. Alternatively, the reward system might simply suggest a new, undiscovered, feed to the user for each unique tag they use. Ultimately this type of reward system could be seen to increase the potential for unsolicited tagging by the users which in turn would provide the stimulus for the future potential of an industry based folksonomy development, e.g. Google Image Labeler, http://images.google.com/imagelabeler/.

7.5 Conclusions

This chapter presented a conceptual proof-of-concept design of a Thoroughbred Industry based Knowledge Portal called "HorseE". The development of a proof-of-concept Paper Prototype was seen as an adequate method for conceptualising and evaluating the concepts of Social Networking and Communal Tagging described in Section 6.4.3, of the previous chapter. The "Wizard of Oz Prototyping" method was used to provide an instructor-led demonstration of the prototype to a number of selected evaluators with interviews with each conducted following the demonstration for the purpose of eliciting feedback.

The feedback received from the evaluation of the HorseE Knowledge Portal prototype proved to be a very constructive, with positive evaluations, negative issues and interesting suggestions being recorded and analysed to provide further insight into the requirements and concerns of the interviewees and potential users. It can be concluded that, although sufficiently easy to use and functional, the proof-of-concept HorseE Knowledge Portal prototype has the prospect for future development, enhancement and extension based on the interview responses.

The HorseE Knowledge Portal is currently a proof-of-concept, but future iterative design, prototyping and evaluation research could be used to bootstrap the development of a Knowledge Sharing initiative, through the use of agile software methodologies, for the Thoroughbred Bloodstock Industry.

8 CONCLUSIONS AND FUTURE WORK

8.1 Introduction

This final chapter summarises and evaluates the findings of the research conducted during the completion of this dissertation. The aim of the research was to examine the current state of learning and Knowledge Management in the Thoroughbred Bloodstock Industry, with a view to proposing an industry-wide Knowledge Sharing initiative, called HorseE. A number of research tasks were performed in the endeavour toward this Knowledge Sharing initiative which included background literature review and culminated in the conducting of a survey to discover the extent of Blog use in the industry and the technical proficiency of its members. Further experimentation and analysis was conducted in an attempt to establish a, currently unspecified, industry specific ontology to be proposed as a Bootstrapping process for community based tagging, categorisation and organisation of Knowledge in the industry. Finally, an experiment was created for evaluation by members of the industry to gauge the acceptability and usability of the initiative concepts identified by the research.

Section 8.2 provides the principal aim of the research and the subsequent research objectives achieved in the dissertation, as well as, a summary of the dissertation chapters. Section 8.3 provides a review of the proposed solution, while section 8.4 provides critical evaluation of the proposed solution based on the experiments carried out and the industry expert feedback obtained. Section 8.5 presents future work and research that has been identified during the process of conducting this research.

8.2 Research Definition and Overview

The ability to share knowledge within the domain of the Thoroughbred Horse Industry has become common practice, by virtue of the accessibility of the World Wide Web resulting in a wealth of knowledge being shared within the Bloodstock Community. Due to a lack of efficient management of this knowledge the domain has become overwhelmed by the extensive amounts of knowledge, with disparate sources of information forming barriers to a collaborative approach, and effective practices for Knowledge Sharing in the Thoroughbred Bloodstock Industry.

The aim of this dissertation has been to explore how Web 2.0 technologies can be effectively leveraged as Knowledge Sharing processes and to evaluate how knowledge is currently shared within the online Bloodstock Community. By examining Social Networking practices and current web semantics and ontologies a proposal of a bootstrapping mechanism for a Social Semantic Knowledge Sharing initiative for the Thoroughbred Bloodstock Industry is presented and evaluated.

8.2.1 Achieved Research Objectives

A number of research objectives were achieved during the course of the research and experimentation in this dissertation which can be seen to contribute to the endeavour toward a Knowledge Sharing initiative for the Thoroughbred Bloodstock Industry. These research objectives have been identified as follows:

- An examination of the current state of Learning and Knowledge Sharing methods in the Thoroughbred Bloodstock Industry was carried out. This examination highlighted deficiencies in the approaches towards Knowledge Sharing by industry members and identified areas for improvement.
- An extensive investigation was conducted to identify how current Web 2.0 technologies, including Social Software and Web Portals, can be leveraged as viable sources of Knowledge, and ultimately viable mechanisms for Knowledge Sharing, in the Thoroughbred Bloodstock Industry.
- An industry-wide survey was conducted to assess the understanding and use of Blogs by members of the Thoroughbred Bloodstock Industry. This survey also served to highlight the level of expertise of the participants and their technical proficiencies with web and blog usage.
- Web 2.0 technologies were leveraged, specifically Blogs and RSS feeds, through the use of an RSS Reader and Feed Aggregator to consume and extract feed content from a cross-sectional representation of the industry knowledge.
- An indepth analysis of the extracted feed content was conducted which facilitated the proposal and development of a basic domain-specific ontology for the Thoroughbred Bloodstock Industry, that can be used to bootstrap the process of Knowledge Sharing.

- A hybrid social semantic environment was proposed, leveraging SIOC and FOAF technologies, to facilitate an industry-wide Community of Practice where knowledge can be shared through the use of Social Networking and Communal Tagging.
- A proof-of-concept prototype was proposed and developed using the "Paper Prototyping" method for user evaluation and testing.
- Finally user evaluation and feedback based on the proof-of-concept prototype was compiled to provide insight into the possible acceptance and usability of the design for the HorseE Knowledge Portal.

8.2.2 Summary of Dissertation

Chapter 1 provided the introduction, background and a description of the research problem being investigated, the main objectives of the research and the deliverables of the project.

Chapter 2 presented the current Thoroughbred Bloodstock Industry as a Learning Industry and challenged the fundamental principles of traditional organisational Knowledge Management by discussing the difficulties involved in developing a Knowledge Sharing initiative in the context of a Knowledge Industry. A solution was presented by virtue of Communities of Practice and more recently, Virtual Communities of Practice.

Chapter 3 explored Web 2.0, its Social Software technologies, and its evolution of the traditional Knowledge Management principles from the formal organisational to more social community based interactive methodologies. Weblogs are presented as a medium for the Acquisition, Packaging, Storing and Sharing of Knowledge within a wider community of web users through the use of RSS sharing/syndication facilities.

Chapter 4 examined the concept of progressive social interaction on the Web through the consumption of RSS feeds by portals and the ability of portal users to categorise and organise knowledge through communal tagging and folksonomies. The concept of formalised categorisation using taxonomies and domain-specific ontologies was

introduced followed by a critique of the use of Portals for Knowledge Sharing in the Thoroughbred Bloodstock.

Chapter 5 presented strong evidence that blogs are a key enabler of Knowledge Sharing within the Thoroughbred Bloodstock Industry, based on an industry survey. Although Knowledge Internalisation is identified as a determining factor for the use of Blogs evidence is presented indicating a definite shift in knowledge management and sharing practices in the Thoroughbred Bloodstock Industry, from Socialisation to Externalisation.

Chapter 6 assessed the suitability of RSS for the conveyance of organised domain knowledge and proposed the process of key term identification to develop an industry folksonomy. A basic domain-specific ontology was developed and a Hybrid Tagging System was proposed to bootstrap the interaction between an industry folksonomy and domain-specific ontology.

Chapter 7 offered a conceptual proof-of-concept prototype of a Thoroughbred Bloodstock Industry based Knowledge Portal for conceptualising and evaluating the concepts of Social Networking and Communal Tagging, identified in Chapter 6.

8.3 Contributions to the Body of Knowledge

The research outlined in this dissertation has achieved a number of contributions to the body of knowledge relating to Knowledge Management and Knowledge Sharing within the Thoroughbred Bloodstock Industry. The reviewed literature indicates that traditional Knowledge Management (KM) focuses on the development of systems used to collect and organise the knowledge within an organisation for the purpose of exposing and making the knowledge more accessible and useful to the organisation. Such KM systems should continuously evolve to incorporate new knowledge in order to generate new knowledge and expose this knowledge in an organised modus that is accessible by people within the organisation. It should be noted that the majority of KM literature is predominantly focused on the concept of KM within organisations and few examples can be found relating to industry-wide KM, primarily within the Thoroughbred Bloodstock Industry.

The author has identified numerous comparable KM concepts which can be exploited and correlated to KM concepts within an Industry-wide environment. However it has been identified that a number of concepts including; Managers, Employees, etc., need to be exchanged for relative Industry terms, such as Expert, Member, Participant, etc., in order to be applicable in the industry domain. Additionally the removal of geographic limitations within an industry are identified as a distinct variation of the concept of culture between an organisation and an industry, which should be given due consideration when reviewing and resolving the barriers to Knowledge Sharing within a Knowledge Industry.

The paradigm shift from a traditional centralised repository to a more conversational collaborative intelligence is, in the author's view, a pivotal catalyst for the evolution of the Thoroughbred Bloodstock Industry from a real Community of Practice to a more socially oriented Virtual Community of Practice, through leveraging existing and emerging social and semantic web technologies. However, the author also recognises that a current lack of technological proficiency with the use of such technologies, for example the use of uninformative Tags on Blog articles and within RSS feeds, in the Thoroughbred Bloodstock Industry, resulting in limitations to the successful implementation of a Knowledge Sharing initiative.

To this end the author proposed an alternative mechanism for identifying key industry terms, used to form an industry folksonomy, through analysis of the article content of syndicated feeds. In the author's view this proved a successful exercise as the emerging folksonomy tended toward agreement with the theory of Power Law distributions, indicating that given time the folksonomy would evolve and eventually depict the key concepts and terms related to the industry.

Through further examination of the emerging folksonomy, development of a basic domain-specific ontology and proposed incorporation of semantic ontologies, namely SIOC and FOAF, the author has successfully visualised the potential for a Hybrid Tagging Ontology. This Hybrid Tagging Ontology which exploits the benefit of both the ease of use of folksonomies and the support of the formalisms and the methods of the Semantic Web, in the author's view, has the impetus for providing new and

improved perspectives for assisting Knowledge Sharing on the "Socially Semantic Web".

8.4 Experimentation, Evaluation and Limitation

Having distributed 300 invitations for participation in an industry survey, only 48 responses were received from a total of 280 delivered invitations. The distribution of invitations included a representative cross-section of 150 Stud Farms, 100 Trainers and 50 Bloodstock Agents from both the Northern and Southern hemispheres. The total number of responses resulted in a response rate of 17% for this research after three reminder invitations; a typical response rate is generally about 20% (Burgess, 2001). According to Yun and Trumbo (2000), response rates may reach 25% to 30% without follow up reminder emails with some email surveys achieving response rates as high as 70. Although the response rate is not unusually low some of the feedback received indicates that there is reluctance on behalf of respondents to reveal their level of knowledge.

Lack of interest, lack of technical knowledge, limited technical resources and limited time are some of the factors identified by virtue of the author's 11 years of working experience servicing the Thoroughbred Bloodstock Industry and significant knowledge of its collective mentality, which have contributed to the low response rate. In a number of respondent cases, limited time was cited as the reason for non-participation in the survey. It is the author's opinion that this type of response is an indication that the labour and time intensive nature of the industry has significantly contributed to the lack of an adequate Knowledge Sharing initiative in the Thoroughbred Bloodstock Industry, as the domain experts do not have the time or resources to promote such an initiative.

The main experiment performed during completion of the project uses word count and frequency distribution metrics for analysis of RSS feed content in an attempt to identify the key terms and concepts of the Thoroughbred Bloodstock Industry. However, these metrics are limited to the syntax level, which has a bias based on the assumption the more words a section contains, the more important that word is. The lack of a suitable weighting system for identified keyword importance was seen as the

main limitation in the experiment, however in the authors view this limitation was acceptable as the intention of the analysis was to produce a basic folksonomy of Industry Terms and not their importance.

It is the author's belief that the implementation of a structured methodology or algorithm for analysing the RSS feed content contributed to successful identification and modelling of relevant key industry terminology. The algorithm implemented consisted of the following steps:

1. Compilation of recognised Stop Words.
2. Identification of the Optimal Starting Point based on word length distribution.
3. Implementation of SQL Queries and Views for presenting the data in a format for analysis.
4. Additional data cleansing and examination to identify incorrect of missing terms. This step sought to counteract the bias introduced based on importance of word frequency and to reintroduce industry terms that were removed through use of standard Stop Words.
5. Further Word Analysis conducted to identify synonyms and participles that could be amalgamated to remove noisy results and outliers.
6. Data Analysis of the resultant folksonomy to confirm the suitability for use and the successful modelling of industry concepts.

The final component of the experiment involved the development and evaluation of a prototype for an industry based Knowledge Sharing initiative by a number of industry related individuals. In the author's opinion the use of "Paper Prototyping", for design, and "Wizard of Oz Prototyping", for evaluation, were essential as they represented a simple yet powerful method for relating the modelled concepts to the evaluators.

In the context of the research the method of evaluation proved significant as the instructor-led process provided the evaluators with time to understand and absorb the concepts of Social Networking and Communal Tagging as they related to the Thoroughbred Bloodstock Industry. The evaluation interviews proceeded by means of interactive conversation, questions and answers, which the author feels produced more

significantly indepth comments (Positive, Negative and Interesting) than if the evaluators were required to test the prototype without guidance.

In the context of this dissertation, portals have been considered to be potential tools for Knowledge Management and for providing mechanisms for industry-wide Knowledge Sharing, however one of the most significantly interesting findings arising from the evaluation is the perceived application of the HorseE Knowledge Portal prototype as a marketing and business tool. This provides a direct correlation with previous research indicating that the motivation behind the development of Information or Knowledge Portals is to attract users to the portal for business purposes. It is therefore the author's belief that the success of any Knowledge Management or Knowledge Sharing initiative within the Thoroughbred Bloodstock Industry must have adequate business and marketing incentive to succeed.

8.5 Future Work & Research

Four main areas of future work and potential research have been identified by the author during the course of this research. One aspect of future works would consist of additional data cleansing processes that could be implemented during the Feed Content Analysis algorithm. Since the Thoroughbred Bloodstock Industry is a very specialist domain, the use of more relevant domain-specific stop words might improve the consistency and relevance of the emergent industry folksonomy.

As indicated in Section 6.5, the familiarity of the Thoroughbred community with descriptive terminology, such as the Names of Horse, Trainers, Stud Farms, Racecourse, etc. indicates that removal of these descriptive terms may destabilise the folksonomy. Therefore careful consideration would need to be given to the introduction of domain-specific stop words and possible indication of when they should be used, e.g. contextual stop words, used in certain domains in the industry.

Chapter 6 also provided an indication for a second area for future work, involving the analysis of industry term variance over time. In the context of this research analysis of this form was not deemed to be significant as the feed analysis was conducted in order to acquire a current snapshot of the key terminology within the. The terminology can

be viewed as fairly static as the industry has been established for more than 250 years, but further examination would be used to support this and also identify if cross-culturalisation and globalisation of the online industry have had an effect on the language used within the industry.

A third area for further work is the iterative development of the HorseE Knowledge Portal prototype. Currently in a proof-of-concept state, the prototype could be further developed to incorporate suggestions and observations elicited during the evaluation process, with a view to liaising with industry experts and users in order to improve and build on the concepts modelled in the existing prototype.

Finally, the process of extracting the collective folksonomy derived from the tags contributed by all users of the HorseE Knowledge Portal and developing and expanding a domain-specific ontology based on this Communal Tagging process is seen as an important element of future work and research derived from this dissertation. This would use the concepts proposed in Section 6.4.3 in order to provide a mechanism through which the collective folksonomy could bootstrap the ontology development. The developed, or extended, domain-specific ontology could in turn be used to further formalise the tags suggested within the HorseE Knowledge Portal, creating a cyclical revision of the folksonomy, ontology and tagging processes.

BIBLIOGRAPHY

Alexander, B., 2006. Web 2.0: A new wave of innovation for teaching and learning? *Educause Review*, 41(2), p.32–44.

Al-Khalifa, H.S. & Davis, H.C., 2007. Towards better understanding of folksonomic patterns. In *Proceedings of the eighteenth conference on Hypertext and hypermedia*. HT '07. New York, NY, USA: ACM, p. 163–166.

Allee, V., 1999. The art and practice of being a revolutionary. *Journal of Knowledge Management*, 3(2), pp.121-132. Available at: [Accessed March 21, 2011].

Andriole, S.J., 2010. Business Impact of Web 2.0 Technologies. *Comunications of the ACM*, 53(12), pp.67-79.

Antonacopoulou, E. et al., 2006. *Learning, Working and Living: Mapping the Terrain of Working Life Learning* annotated edition., Palgrave Macmillan.

Ardichvili, A., Page, V. & Wentling, T., 2003. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, 7(1), pp.64-77. Available at: [Accessed March 20, 2011].

Avram, G., 2006. At the Crossroads of Knowledge Management and Soical Software. *The Electronic Journal of Knowledge Management*, 4(1), pp.1-10.

Awad, E.M. & Ghaziri, H.M., 2003. *Knowledge Management* United States ed., Prentice Hall.

Berners-Lee, T., Hendler, J. & Lassila, O., 2001. The Semantic Web. *Scientific American (Feature Article)*.

Bernsen, N.O., Dybkjær, H. & Dybkjær, L., 1993. Wizard of Oz prototyping: How and when. *CCI WORKING PAPERS IN COGNITIVE SCIENCE AND HCI*. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.135.2044.

Bertland, L., 2001. Internet Portals. Available at: http://www.sldirectory.com/searchf/portals.html#top.

Bojars, U. et al., 2008. Interlinking the Social Web with Semantics. *IEEE Intelligent Systems*, May/June 2008, pp.29-40.

Bojars, U. et al., 2010. The SIOC Project: Semantically-Interlinked Online Communities.

Boyd, S., 2006. Are You Ready For Social Software? *Message*. Available at: http://www.stoweboyd.com/message/2006/10/are_you_ready_f.html.

Brannigan, N., 2010. Enhancing Leadership Capacity in ICTs in Education through technology enabled collaboration. Available at: http://www.gesci.org/assets/files/Enhancing%20Leadership%20capacity%20in%20ICTs%20in%20Education%20through%20Technology%20enabled%20Collaboration%281%29.pdf.

Brown, J.S. & Duguid, P., 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science*, 2(1), p.40–57.

Bryant, L., 2003. *Smarter, Simpler, Social*, Available at: www.headshift.com.

Burgess, T.F., 2001. A general introduction to the design of questionnaires for survey research.

Byrne, G. & Goddard, L., 2010. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine*, 16(11/12). Available at: http://www.dlib.org/dlib/november10/byrne/11byrne.html [Accessed March 21, 2011].

Cayzer, S., 2004. Semantic Blogging and Decentralized Knowledge Management. *Communications of the ACM*, 47(12), pp.47-52.

Chen, R.-C., Bau, C.-T. & Yeh, C.-J., 2011. Merging domain ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques. *Applied Soft Computing*, 11(2), pp.1908-1923. Available at: [Accessed March 21, 2011].

Coates, T., 2003. My working definition of social software... *PlasticBag.org*. Available at: http://www.plasticbag.org/archives/2003/05/my_working_definition_of_social_software/.

Collison, C. & Parcell, G., 2002. *Learning to Fly*, Capstone: London.

Cormode, G. & Krishnamurthy, B., 2008. *Key Differences between Web 1.0 and Web 2.0*, At&T Labs-Research.

Dash, A., 2002. Introducing the Microcontent Client. Available at: http://dashes.com/anil/2002/11/introducing-microcontent-client.html.

Després, C. & Chauvel, D., 2000. *Knowledge horizons: the present and the promise of knowledge management*, Butterworth-Heinemann.

Downes, S., 2005. E-learning 2.0. *ACM eLearn Magazine*.

Dukes, A., 2009. *Analysis of the economic impact of the Irish Thoroughbred Horse Industry*, Irish Thoroughbred Breeders' Association.

Efimova, L., 2004. Discovering the Iceberg of knowledge work: A weblog case. In European Conference on Organizational Knowledge, Learning, and Capabilities OKLC 2004.

Ethridge, R.R., Hadden, C.M. & Smith, M.P., 2000. Building a Personalized Education Portal: Get a Behind-the-Scenes Look at LSU's Award-Winning System. *Educause Quarterly*, 23(3), pp.12-19. Available at: [Accessed March 20, 2011].

Fahey, L. & De Long, D.W., 2000. Diagnosing cultural barriers to knowledge management. *Academy of Management Executive*, 14(4), pp.113-127.

Field, L. & Ford, G.W. (Gordon W., Melbourne : Longman Australia. *Managing organisational learning : from rhetoric to reality / Laurie Field in collaboration with Bill Ford*, Melbourne : Longman Australia.

Financial Times Lexicon, 2011. knowledge industry definition from Financial Times Lexicon. Available at: http://lexicon.ft.com/Term?term=knowledge-industry [Accessed March 21, 2011].

Godwin-Jones, R., 2003. *Emerging technologies: blogs and wikis: environments for on-line collaboration. : An article from: Language, Learning & Technology*, Available at: http://www.amazon.com/exe%bidos/redirect?tag=citeulike07-20&path=ASIN/B00082CIDC.

Goetz, T., 2003. Open Source Everywhere. *Wired*. Available at: http://www.wired.com/wired/archive/11.11/opensource_pr.html.

Gross, T. et al., 2009. *Human-Computer Interaction - INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedigns*, Springer.

Gruber, Thomas, 2007. Ontology of Folksonomy. *International Journal on Semantic Web and Information Systems*, 3(1), pp.1-11. Available at: [Accessed March 20, 2011].
Gruber, Tom, 1992. What is an Ontology? Available at: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

Halpin, H., Robu, V. & Shepherd, H., 2007. The Complex Dynamics of Collaborative Tagging. In *E* Track*. 16th International World Wide Web Conference. Banff, Alberta, Canada.

Halvey, M.J. & Keane, M.T., 2007. An assessment of tag presentation techniques. In *Proceedings of the 16th international conference on World Wide Web*. WWW '07. New York, NY, USA: ACM, p. 1313–1314.

Han, S., 2003. *Semantic Web Community Portal Project*,

Hansen, M.T., Nohria, N. & Tierney, T., 2000. *What's Your Strategy for Managing Knowledge?*, Harvard Business Review.

Harrsch, M., 2003. RSS: The Next Killer App For Education. Available at: http://technologysource.org/article/rss/.

horseracing.com, 2011. The History of Horse Racing. Available at: http://www.horseracing.com/history-of-horse-racing/.

Hotho, A. et al., 2006. Information Retrieval in Folksonomies: Search and Ranking. *The Semantic Web: Research and Applications*, 4011, pp.411-426.

Hurwitz, J.S., 2001. Portals Tie Problems To Solutions. *InformationWeek*, (820), p.110.

InContext, 2011. Contextual Design. *incontext Customer Centered Design*. Available at: http://incontextdesign.com/contextual-design/.

Kimble, C., Hildreth, P. & Wright, P., 2001. *Communities of Practice: Going Virtual*, Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.5217.

Kirchner, K., Razmerita, L. & Sudzina, F., 2009. New Forms of Interaction & Knowledge Sharing on Web 2.0. *Web 2.0*, pp.1-16.

Klamma, R., Cao, Y. & Spaniol, M., 2007. Watching the Blogosphere: Knowledge Sharing in the Web 2.0. In International Conference on Weblogs and Social Media. Boulder, Colorado, USA.

Knerr, T., 2007. Tagging Ontology - Towards a Common Ontology for Folksonomies. Available at: http://code.google.com/p/tagont/downloads/detail?name=TagOntPaper.pdf.

Kreuder, A.C. & Schiefer, G., 2002. INFRASTRUCTURE OF INFORMATION-PORTALS FOR THE AGRIFOOD SECTOR – INFORMATION DEDUCTION AND PORTAL MANAGEMENT. Available at: http://uf.ilb.uni-bonn.de/uf/publika/pdfVeroeffilb/efita03_kreuder.pdf.

Krishnamurthy, S., 2002. The multidimensionality of blog conversations: The virtual enactment of september 11. In *Internet Research 3.0*.

Lamantia, J., 34. Creating Successful Portals with a Design Framework. *International Journal of Web Portals*, 1(4), pp.63-75. Available at: [Accessed March 20, 2011].

Lave, J. & Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation* 1st ed., Cambridge University Press.

Lee, M.R. & Lan, Y.-chen, 2007. From Web 2.0 to Conversational Knowledge Management: Towards Collaborative Intelligence. *Journal of Entrepeneurship Research*, 2(2), pp.47-62.

Levy, M., 2009. WEB 2.0 implications on knowledge management. *Journal of Knowledge Manaement*, 13(1), pp.120-134.

Limpens, F., Gandon, F. & Buffa, M., 2008. Bridging Ontologies and Folksonomies to Leverage Knowledge Sharing on the Social Web: a brief survey. In The 1st International Workshop on Social Software Engineering and Applications. L'Aquila, Italy.

Liu, J., Wong, C.K. & Hui, K.K., 2003. An Adaptive User Interface Based On Personalized Learning. *IEEE Intelligent Systems*, 18, p.52–57. Available at: [Accessed March 21, 2011].

Lohmann, S., Ziegler, J. & Tetzlaff, L., 2009. Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*. INTERACT '09. Berlin, Heidelberg: Springer-Verlag, p. 392–404. Available at: http://dx.doi.org/10.1007/978-3-642-03655-2_43 [Accessed March 20, 2011].

Long, D.W.D. & Fahey, Liam, 2000. Diagnosing Cultural Barriers to Knowledge Management. *The Academy of Management Executive (1993-2005)*, 14(4), pp.113-127. Available at: [Accessed March 20, 2011].

MacManus, R. & Porter, J., 2005. Web 2.0 Design: Bootstrapping the Social Web. *Digital Web Magazine*. Available at: http://www.digital-web.com/articles/web_2_for_designers/.

Marsick, V.J. & Watkins, K.E., 1999. *Facilitating learning organizations: making learning count*, Gower Publishing, Ltd.

Mejias, U., 2005. A Nomad's Guide to Learning and Social Software. Available at: http://knowledgetree.flexiblelearning.net.au/edition07/html/la_mejias.html.

Mejias, U., 2006. Teaching Social Software with Social Software. *innovate*, 2(5).

Merriam-Webster A, A., 2010. Definition of Knowledge. Available at: http://www.merriam-webster.com/dictionary/knowledge.

Merriam-Webster B, B., 2010. Definition of Culture. Available at: http://www.merriam-webster.com/dictionary/culture?show=0&t=1283375082.

Mikroyannidis, A., 2007. Toward a Social Semantic Web. *Computer*, 40(11), pp.113-115.

Morel, E. & Smith, D.P., 2010. Courses About Horses – Graduates in the Horse Industry. *Racetrack Magazine*, pp.24-26.

Morrison, C., 2000. *Knowledge and Learning in the Horse Industry*, Australia: Rural Industries Research and Development Corporation.

Mortensen, T. & Walker, J., 2002. Blogging thoughts: personal publication as an online research tool. In *Researching ICTs in context*. Oslo: InterMedia/UniPub, pp. 249-279.

Niles, R., 2005. Wikis will help readers direct the community's most powerful voice: opinion writers from younger generation offer fresh perspective. *The Masthead*, (Autumn).

Nonaka, I. & Takeuchi, H., 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, USA.

Noy, N.F. & McGuinness, D.L., 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*, Available at: http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html.

O'Leary, M., 2000. Grading the library portals. *Online*, November 2000. Available at: http://www.onlinemag.net/OL2000/oleary11.html.

O'Reilly, 2006. *Web 2.0 Principles and Best Practices*, O'Reilly Media.

O'Reilly, T., 2005a. Web 2.0: Compact Definition? Available at: http://radar.oreilly.com/archives/2005/10/web-20-compact-definition.html.

O'Reilly, T., 2005b. What Is Web 2.0'.

Ojala, M., 2004. Weaving Weblogs into Knowledge Sharing and Dissemination. In *Knowledge and Change*. Nord I&D. Denmark.

Oravec, J.A., 2004. Time Capsules: A Cultural History. *The Journal of Popular Culture*, 37(4), pp.735-737. Available at: [Accessed March 21, 2011].

Park, J., Oh, S. & Ahn, J., 2011. Ontology selection ranking model for knowledge reuse. *Expert Systems with Applications*, 38(5), pp.5133-5144. Available at: [Accessed March 20, 2011].

Passant, A., 2007. Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs: Theoretical background and corporate use-case. In International Conference on Weblogs and Social Media. Bouldar, Colorado, USA.

Peek, R., 1999. When Is a Web Site a "Vortal"? *Information Today*, September 1999. Available at: http://www.highbeam.com/doc/1G1-55676554.html.

Pitner, T. et al., 2007. Web 2.0 as Platform for Inclusive Universal Access in Cooperative Learning and Knowledge Sharing. In *Proceedings of I-Know '07*. I-Know '07. Austria.

Polanyi, M., 1983. *Tacit Dimension*, Peter Smith Publisher Inc.

Pouchard, L., Ivezic, N. & Schlenoff, C., 2000. Ontology Engineering for Distributed Collaboration in Manufacturing. In *Proceedings of the AIS2000 conference*. AIS2000.

Powers, S. et al., 2002. *Essential Blogging: Selecting and Using Weblog Tools* 1st ed., O'Reilly Media.

Pressley, L., 2006. Using Social Software for Business Comminication. Available at: http://laurenpressley.com/papers/socialsoftware_business.pdf.

Riege, A., 2005. Three-dozen knowledge-sharing barriers managers must consider. *Journal of Knowledge Management*, 9(3), pp.18-35.

Rollett, H. et al., 2007. The Web 2.0 way of learning with technologies. *International Journal of Learning Technology*, 3(1), pp.87-107.

Rosch, E. et al., 1976. Basic objects in natural categories. *Cognitive Psychology*, 8(3), pp.382-439. Available at: [Accessed March 21, 2011].

Rowley, J., 2000. Portal Power. *Managing Information*, 7(1), pp.62-64.

Salari, J., 2009. tag, you're it. Available at: http://www.cs.ubc.ca/~jamila/TaggingPaper.pdf.

Schechter, S., Egelman, S. & Reeder, R.W., 2009. It's not what you know, but who you know: a social approach to last-resort authentication. In *Proceedings of the 27th international conference on Human factors in computing systems*. CHI '09. New York, NY, USA: ACM, p. 1983–1992.

Senge, P.M. et al., 1999. *The Dance of Change: The Challenges to Sustaining Momentum in Learning Organizations* 1st ed., Crown Business.

Snyder, C., 2001. Paper Prototyping. *IBM developerWorks*. Available at: http://snyderconsulting.net/article_paperprototyping.htm.

Stephenson, K., 2004. What knowledge tears apart, networks make whole. *Internal Communication Focus*, No. 36.

Stewart, T.A., 1998a. *Intellectual Capital: The New Wealth of Organizations* 1st ed., Crown Business.

Stewart, T.A., 1998b. *Intellectual Capital: The New Wealth of Organizations* 1st ed., Crown Business.

Straker, D., 1997. *Rapid Problem Solving with Post-It Notes*, Da Capo Press.

Technorati, 2009. *State of the Blogosphere 2009*, Available at: http://technorati.com/blogging/feature/state-of-the-blogosphere-2009/.

Tu, C.H., 2002. The Measurement of Social Presence in an Online Learning Environment. *International Journal on E-Learning*, 1(2), p.34–45.

W3C, 2011. About The World Wide Web. Available at: http://www.w3.org/WWW/.

Wagner, C., 2005. Breaking the Knowledge Acquisition Bottleneck through Conversational Knowledge Management. *Information Resources Management Journal*, 19(1), p.70–83.

Weber, R.P., 1990. *Basic Content Analysis* 2nd ed., Sage Publications, Inc.

Weller, K., 2007. Folksonomies and ontologies: two new players in indexing and knowledge representation. In *Online Information 2007 Proceedings*. Online Information 2007.

Wenger, E., 2006. Communities of practice. Available at: http://www.ewenger.com/theory/index.htm [Accessed March 21, 2011].

Wenger, E., White, N. & Smith, J.D., 2009. *Digital Habitats; stewarding technology for communities*, CPsquare.

Wikipedia A, 2011. Knowledge sharing - Wikipedia, the free encyclopaedia. Available at: http://en.wikipedia.org/wiki/Knowledge_sharing [Accessed March 21, 2011].

Wikipedia B, 2011. Tag (metadata) - Wikipedia, the free encyclopaedia. Available at: http://en.wikipedia.org/wiki/Tag_(metadata)#Within_a_blog [Accessed March 21, 2011].

Wikipedia C, 2011. Ontology (information science) - Wikipedia, the free encyclopaedia. Available at:

http://en.wikipedia.org/wiki/Ontology_(information_science) [Accessed March 21, 2011].

Yun, G.W., 2000. Comparative Response to a Survey Executed by Post, E-mail, & Web Form. *Journal of Computer-Mediated Communication*, 6(1), p.2000. Available at: [Accessed March 20, 2011].

Zárraga-Oberty, C. & Saá-Pérez, P.D., 2006a. Work teams to favor knowledge management: towards communities of practice. *European Business Review*, 18(1), pp.60-76. Available at: [Accessed March 20, 2011].

Zárraga-Oberty, C. & Saá-Pérez, P.D., 2006b. Work teams to favor knowledge management: towards communities of practice. *European Business Review*, 18(1), pp.60-76. Available at: [Accessed March 21, 2011].

Zhou, L., Booker, Q. & Zhang, D., 2002. Toward Rapid Ontology Development for Underdeveloped Domains. In *Hawaii International Conference on System Sciences*. Los Alamitos, CA, USA: IEEE Computer Society, p. 106.