KNOWLEDGE MEDIA INSTITUTE

The Modelling, Capture, and Use of Social Context in Online Tasks

Technical Report kmi-05-10
September 2005

Tom Heath

The Open University

Table of contents

Abstract

This report consists of three parts. Part I reviews how users' online tasks have been conceptualised in previous literature, and how researchers have defined and used context in support of user tasks. Novel conceptualisations of user tasks online and user context factors are then presented and contrasted with earlier work, before a discussion of how these context factors have been supported in previous applications. The modelling of social context is then considered in greater detail, with particular focus on aspects such as the nature of social relationships and trust between individuals. Research gaps identified through this review of the literature are summarised to conclude this section. Part II addresses specific outputs of the research to date. In particular, the conceptualisations of user tasks online and user contexts are discussed in more detail, including coverage of the assumptions they are based upon and the background to their development. Specific technical work carried out is also described, including the planning of a social context application, analysis of tools and technologies that may be utilised, and development of relevant technical skills. Drawing on the gaps identified in Part I of the report, Part III introduces the questions that will be addressed by the research. After justifying the research questions, the methods that will be used are outlined and discussed, including overall plans for how the research will be carried out.

Introduction

The literature reviewed in Part I of this report highlights two problems that underpin the research discussed here. Firstly, current internet applications do not adequately support users performing tasks online, and secondly do not take the user's context into account to assist in performing these tasks.

A review of current conceptualisations of the tasks users perform online suggests that these are limited by the following factors: a focus on individual internet platforms (such as the web) at the expense of a holistic view; confusion of the user's needs with strategies they use and with artefacts of current internet applications; assumptions of a search- or browse-centric mode of interaction; and by little relation to the tasks performed and strategies used in the offline world. This reveals a theme whereby specific application-level internet protocols rather than the goals of the user define the tools available to them. In turn use of these tools shapes the conceptualisations we hold of the tasks users perform online.

Similarly, a review of how context is currently utilised in support of online tasks reveals wide variety in how context is defined, and a number of attempts to support specific aspects of it such as the context of search queries and the computing resources available. These factors demonstrate that context is not perceived or supported in a unified way and is frequently under utilised in support of user tasks online.

In response to these perceived limitations, Part II of this report presents two conceptualisations as the basis for user-oriented internet applications. Firstly, a unified view of the tasks users perform online is presented that focuses on the goals of the user in being online without assuming the use of particular tools or strategies in achieving these goals. Secondly, a number of user context factors are introduced that are seen to influence and assist with online tasks as they might tasks in the offline world.

Looking specifically at the social context factor raises a number of questions that form the basis of the research proposed in this report. At the highest level this takes the form: can a person's social context be modelled, captured, and utilised to facilitate locating and exploring information online? At a more granular level, this raises questions regarding how people use their social networks for locating and exploring information, whether emerging semantic web technologies can enable the modelling, capture and use of this social context, and whether such elements can be integrated into a system that is both useful and usable.

Addressing these questions and realising a more contextualised online experience has real value in addressing the user need of reducing information overload, a problem frequently experienced at present when locating and exploring information online. Furthermore it has the potential to reduce the distinction between offline and online worlds, by enabling task performance strategies that exist in the offline world to be more easily utilised in the online world. Finally, research into the modelling, capture, and use of social context could inform further research into how the same may be achieved with the other context factors introduced in this report.

Part III of this report outlines three research components to be carried out to address these questions. Firstly, a user study is proposed to further investigate how social networks are used in locating and exploring information. Specifically, this will examine the role of factors such as the nature of the social relationship, as well as the perceived domain expertise and proximity of others within the network, in determining how people use their social network to help them perform these tasks. The second component consists of research into modelling social context, and developing an application capable of capturing and utilising this in support of online tasks, using semantic web technologies. The final component of the research consists of evaluation of this application to assess: the feasibility of developing such an application using semantic web technologies; its validity in supporting the strategies reported in the user study; and its utility relative to current online tools and offline approaches.

Part I. Literature review

Introduction

Many everyday tasks, such as shopping, banking, education, finding holiday and travel information, reading newspapers and magazines, and communicating can now be carried out online and regularly are by significant numbers of people (Emmanouilides and Hammond, 2000). Throughout the history of the internet, numerous tools and protocols have been developed that form distinct platforms and define the online experience users are familiar with. These platforms include, for example, email, the web, ftp, newsgroups, mailing lists (Emmanouilides and Hammond, 2000) and are regularly used in performing the kinds of tasks outlined above.

It is likely that the average user is concerned with the task they went online to perform, rather than the internet's underlying protocols. However, these platforms often delineate studies of internet usage, with researchers focusing on use of a particular protocol or application, in the process hiding the true goal of the user in being online. For example, Sellen, Murphy and Shaw (2002) investigate the tasks that knowledge workers perform on the web, but exclude web-based email access from their analysis on the basis that this is a different type of application. This application- and protocol-centric view can be seen to both constrain ideas about the functions a platform serves, and constrain users within the boundaries of specific applications that may not be fully suited to their needs.

For example, the web was originally designed to provide access to information and allow links to be created between documents. As use of the web increased, the underlying technologies were complimented by others that enabled more interactivity, leading to the development of *webtop* applications such as banking and email. Some of these applications operate on the server side with users being sent the results for display in a web browser, whilst others use the browser itself as the execution environment. These factors indicate a fundamental change in how the web and the internet as a whole are being used.

A comparable process can be observed in how email use has evolved over time. Beyond the original purpose of allowing people to exchange messages based on a memo-style format, email is now commonly used for tasks as diverse as discussion, notifying others, file sharing and storage, and task or project management.

Despite these changing use patterns, many of the internet tools used, and the ways of thinking about user behaviour online, remain rooted in a protocol-centric perspective. In this context we need to develop a more abstract understanding of the tasks users do or want to perform online. This understanding needs to occupy a middle ground between seeing online tasks in very specific terms, such as banking or reading a newspaper on the web, and the protocol- or application-specific view of what the internet is used for.

The semantic web vision extends the idea of the web from an information retrieval-centric platform to one of knowledge, tasks, agents, and users. The seminal semantic web paper (Berners-Lee, Hendler, and Lassila, 2001) features two characters Pete and Lucy, their agents, and several tasks that must be performed. For example, Lucy says to Pete: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Bi-weekly or something. I'm going to have my agent set up the appointments" (pp. 36). This is an example the web being used for arranging something, rather than simply as a source of information.

The opportunity exists to extend semantic web ideas across all internet protocols and platforms and make an online experience for the user that is less detached from the offline world. To do this we need to understand the context of the user as they perform tasks online,

and this should be informed by how tasks are performed offline, not by a technology driven view.

This section of the report will review conceptualisations of the tasks people perform online, ways that previous authors have characterised context, and how context has been used to support performance of user tasks online. The review will conclude with a specific focus on social context, discussion of how it can be better understood and used to support tasks online, and a summary of the gaps in existing research.

Internet usage as task performance

For the purpose of this report, a task is defined as "an activity or action that has a specific goal or intention". This goal or intention does not have to result in a specific tangible consequence or outcome, as it is recognised that the task may be rewarding in its own right.

To assess how well existing tools support users in completing tasks online, and how they might be better supported, it is important to understand the types of tasks people perform on the internet. The majority of literature in this area focuses specifically on the medium of the web rather than the internet as a whole. The findings of these studies will be discussed here, whilst the limitations of this approach are discussed in more detail below.

Conventional views of task performance on the web

Whilst the tasks performed on the web have been categorised and conceptualised in a number of different ways, one common theme is the perception of searching and browsing as the dominant modes of interaction (Olston and Chi, 2003). This perception is understandable given the ubiquity of the search engine and the web browser in web usage.

Application-specific interaction

Cockburn and McKenzie, (2000) report a study of user actions at the web browser. Focusing primarily on analysis of the user's *click-stream* and manipulations of the browser, the study gives insights into how people use navigation devices such as the *Back* button, as well as bookmarking and history tools. Unfortunately, approaching the question from the perspective of the point and click interaction paradigm it does not address the issue of the tasks or goals of the user in being online.

Similarly, Byrne, John, Wehrle, and Crow (1999) investigate user actions with a browser. The use of a video study provides naturalistic evidence of user interaction online compared to studies that define the tasks for users to perform. However, analysis covers a mix of user goals and tasks (such as *locate on page: something "interesting"*) with interface and interaction tasks (*configure browser: change cache size*). Furthermore, tasks such as locate on page refer only to the user interactions with a page they have already loaded, not to the user's overall task on the web as a whole.

Web activities as forms of searching and browsing

Previous research has sought to identify and classify user behaviours on the web, mainly by identifying specific modes of searching or browsing. At the most basic level Guha, McCool and Miller (2003) distinguish between navigational and research searches. In a navigational search "the user is using the search engine as a navigation tool to navigate to a particular intended document", whereas a research search is characterised by the user "trying to locate a number of documents which together will give him/her the information s/he is trying to find" (pp. 702).

Broder (2002) describes a taxonomy consisting of three types of web search: navigational, informational, and transactional. The navigational and informational types map closely onto the navigational and research searches proposed by Guha et al (2003), with transactional searches consisting of queries where the user intends "to reach a site where further interaction will happen" (pp. 6), such as a shopping site or a site where images or music can be downloaded. However, the range of possible transactions a user may wish to perform, and the underlying reason for wishing to perform them is not explored. Similarly, consideration is not given to the reason why the user wishes to locate a particular web site. Presumably visiting the site is not an end in itself, but part of the strategy for performing another task such as finding a phone number or arranging car rental.

Related work by Rose and Levinson (2004) yielded top-level categories with many similarities to those of Broder (2002), but also a number of more detailed sub-categories such as download, entertainment, interact, and obtain. Despite a number of examples being given to illustrate these sub-categories, the distinctions between them are often based on technical aspects of how the target object will be used, rather than the nature of the task the user is performing. For example, the target of the download goal is "a resource that must be on my computer or other device to be useful" (pp. 15). The authors give the example of a piece of software; however the same definition could equally apply to the adult movie example used to illustrate the entertainment sub-category. In both cases it appears the user is trying to locate something that they can then make use of irrespective of how this is done.

One common factor in these studies is a search- or browse-centric perspective on web use. These "two predominant interface modes" (Olston and Chi, 2003, pp 177-178) are often taken as the window through which to study user actions on the web. However, such a perspective may prevent a real understanding of the user's goals in being online.

The focus on classifying search behaviours means none of the schemes discussed so far can account for more complex tasks. Whilst the resource-interact goal of Rose and Levinson (2004) and the transactional queries of Broder (2002) suggest an intention to carry out further interaction beyond the search (perhaps indicating a greater overall goal), the search itself is still seen as the user's primary task. No mention is given of, for example, arranging a holiday as an overarching reason for being online, or even for carrying out a search, and whilst no queries such as "arrange holiday" are reported, this likely reflects that users are aware of the narrow scope of search engines and therefore do not enter such queries, rather than a lack of desire to perform such tasks.

Distinguishing between needs and strategies

Drawing on work in domains such as organisation science Choo, Detlor, and Turnbull (1999) highlight a distinction between a user's information needs and the information seeking strategies they employ to meet these needs. A similar distinction could also be made between the task a user intends to carry out online, and the strategies they use to complete this task.

Morrison, Pirolli, and Card (2001) describe a taxonomy of web activities with three variables: the purpose of a search, the method used, and the content of the information being searched for. Whilst these variables appear neatly defined, the classification of some activities suggests the variables may not be mutually exclusive in the form proposed by the authors. For example, some methods are seen to be triggered by a particular goal (find, collect) whereas others are not (explore, monitor). In this case it may be that explore and monitor actually represent goals in their own right, and should be classed under purpose.

Sellen et al (2002) describe a classification that identifies six activities carried out on the web (finding, information gathering, browsing, transacting, communicating, housekeeping), based on a study of web use by twenty-four knowledge workers. This classification is not limited to describing variations of searching or browsing, and does attempt to capture the user's needs or goals in using the web. However, by focusing purely on web-based tasks (excluding communication by email, for example), the classification does make assumptions about the strategies being used in performing tasks online.

Limitations of conventional views of task performance

The literature outlined above demonstrates that there are many ways to conceptualise the activities people perform on the web. But to what extent do these classifications represent a valid account of users' goals when online? In general, the classifications address just a small selection of the tasks users may wish to perform online, they characterise component parts of much larger tasks which are not identified or accounted for, or draw distinctions between tasks where these may not actually exist. By taking a search-centric view of web usage some classifications also make assumptions about the strategies a user might employ. Even some schemes that attempt to distinguish needs from strategies remain driven by the principle of an information need and information seeking strategy, rather than a task need and a strategy for performing it.

These factors suggest that a fuller understanding of the range and nature of tasks performed online is necessary. In contrast to current classifications, any broader conceptualisation must adequately account for a full range of tasks beyond simply web search, and must not assume the use of specific technology such as search engines or web browsers. In fact, rather than focusing solely on the web as the medium, the only assumption made should be of the user performing tasks using an internet connected device. Distinguishing the web from the rest of the internet in the case of task performance would be to confuse the task need with the strategy employed

A unified view of task performance online

Heath, Dzbor, and Motta (2005) present a user-oriented classification of online tasks, reproduced in Part II of this report. The work builds on and extends the work discussed above, such as Sellen et al (2002), and is based on a number of assumptions also detailed in Part II.

In the light of this classification, can existing internet tools be considered fit for purpose? Do they adequately support the tasks people perform online? Some existing tools address the needs of these tasks fairly well. For example, software that reads news feeds from multiple web sites and aggregates the results on a user's desktop are a successful and widely used means of monitoring many sources at once. Unfortunately a similar level of uptake has not been seen with tools that monitor multiple email accounts, perhaps due to a lack of standardised ways of accessing web-based email accounts, and users often have to perform this task manually.

In many circumstances traditional search engines are an effective means of locating objects or information, although they are generally only able to find results based on matching keywords syntactically, and searches are largely limited to textual content due to the complexity of indexing other media such as images or music.

A number of question answering engines such as Ask Jeeves[1] are available that may be able to help evaluate if a certain piece of information is true, although the user may not be sure whether to trust the source of the answer. Furthermore, many comparison web sites exist that are able to evaluate the cheapest place to buy a product, or the fastest route between two points, but they are only able to use information explicitly represented in their databases, rather than reasoning about alternatives that may meet the user's criteria.

In conclusion, some support for these user tasks is available, but it remains limited and solutions are largely confined to specific internet protocols (e.g. HTTP, SMTP/POP) or platforms (e.g. web, email). The result of this is that the user must coordinate a number of applications to perform complex tasks. Heath, Dzbor et al (2005) suggest that the semantic

[1] http://www.ask.co.uk

web may be able to contribute to supporting task performance online, an issue which is discussed below.

Contribution of the semantic web to task performance online

Overview of the semantic web

The Semantic Web vision (Berners-Lee, 1998), (Berners-Lee et al, 2001) proposes an extension of the current web that takes it from a collection of interlinked documents for human consumption to a space where information is sufficiently structured, and the rules that define this structure sufficiently explicit, as to allow machines to understand and reason with it. These basic building blocks of knowledge represented using the Resource Description Framework (RDF), and rules (in the form of ontologies) that facilitate logical reasoning, have been deployed to varying degrees in the tools discussed below.

A seminal paper by Berners-Lee et al (2001) presents a convincing vision of how the semantic web could support users in performing tasks on a web that goes beyond the hypertext model of linked documents. This vision has raised a number of questions and alternative perspectives from other researchers, resulting in a variety of views about the nature of the semantic web.

Marshall and Shipman (2003) outline three contrasting visions of the semantic web, described as *Taming the Web*, *Knowledge Navigator*, and *Federated Knowledge Base*. The Taming the Web perspective sees the semantic web as an application of reference library principles to the web, requiring large amounts of quality metadata and agreed schemas to describe the collection accurately, and supporting people as the end users. In contrast, the Knowledge Navigator vision is of the semantic web as a space occupied by agents capable of inference and reasoning, performing tasks on behalf of people. The challenge of reasoning across different domains is seen as a limitation of this view of the semantic web. Finally the Federated Knowledge Base perspective sees the semantic web as enabling the interoperability of specific applications or databases through the use of formal representations shared by the parties involved.

Whilst these three views contrast significantly with each other, none of them refute the possibility of the semantic web as a platform for supporting users in performing tasks online. In particular, the Taming the Web perspective is heavily aligned with tasks such as locating and exploring, and the Knowledge Navigator view leaves open the possibility of a wide range of tasks being performed. Being concerned more with application to application interaction, the Federated Knowledge Base perspective is neutral with respect to user tasks, although interoperable applications would undoubtedly be required to support them.

Focusing very clearly on the significance of tasks, McBride (2002) states that “the Semantic Web is about creating an infrastructure in which information from a variety of sources can be integrated on demand to achieve some task”. Whilst authors such as Kalfoglou, Harith, Schorlemmer, and Walton (2004) question the focus purely on information integration, the goal of supporting tasks remains valid.

A considerable amount of semantic web research to date has focused on creation of the infrastructure, and this is now being complemented by an increase in user-oriented semantic web research. The semantic web vision opens the possibility of a much more interactive web that better supports users in performing tasks, a direction which is in line with McBride's (2002) suggestion that promoting practical applications is one key to the success of the semantic web. Despite the variety of views on the nature of the semantic web, a focus on supporting user tasks transcends these, as the goals of the user remains the same, irrespective of the ability of the technology to support them or not.

Existing semantic web tools and their support of user tasks

Guha et al (2003) describe a system known as TAP, which seeks to support what they term *research searches*, which could be seen as a means of performing the *exploring* task introduced above. This is achieved by using Semantic Web data to complement search results obtained using conventional information retrieval methods; in this case a conventional search engine. This principle is illustrated with the example of a search for the musician Yo-Yo Ma that returns "his current concert schedule, his music albums, his image, etc." (pp. 702) as well as results obtained from a regular search engine.

Also supporting users in exploring concepts or entities is the browsing tool Magpie (Domingue, Dzbor and Motta, 2004). In contrast to TAP this tool assumes that the user has been able to reach a document that contains some concepts of interest. An ontological layer over the original document then allows the user to invoke semantic services related to that concept. This serves the purpose of providing related information that may not be explicitly mentioned on the page being viewed.

Another tool that builds on the browser metaphor and applies it to the Semantic Web is Haystack (Quan and Karger, 2004). Here the user is able to browse arbitrary collections of RDF metadata through a point and click interface, with links being made between semantically related items. Crucially this tool is able to gather information on a particular topic from multiple sources and assemble it in one place, in contrast to conventional models of web browsing where the user may have to visit several different pages or sites. Although implemented differently (on the web rather than on the desktop), the application CS AKTive Space (Schraefel, Shadbolt, Gibbins, Harris, and Glaser, 2004) provides a similar ability to explore relations between concepts or entities, although in this case the system is limited to the domain of computer science research in the UK.

One feature all these tools have in common is the ability to present the user with new pieces of information, or make new connections between concepts that might not otherwise have been apparent; this ability is a key feature of the Semantic Web. To this end these tools make a significant step towards supporting users in the task of exploring concepts or entities to gain additional knowledge or understanding. However, many of the other tasks identified by Heath, Dzbor et al (2005) such as *arranging* or *evaluating* are not so well supported, and these tools are not able to provide support in these areas. It would seem that Semantic Web technologies can enable more powerful tools that are better able to support user tasks; however, the range of these tools must be extended to cover the full range of tasks users perform online.

Ecological validity of existing tools

In addition to the need for task-focused tools that support the goals of the user, wherever possible the tools should allow or afford the use of strategies that are used in task performance in the offline world, rather than require entirely new strategies for online purposes. Such tools could be seen as being more *ecologically valid*. In experimental psychology, ecological validity is defined as "the validity that a principle discovered in a laboratory setting has outside of that setting, in the field, in the real world" (Reber, 1997). In this case the term could be used to describe tools or strategies that have relevance and applicability in the wider world rather than only in a specific artificial environment or usage scenario.

One area where increased ecological validity could be achieved is in creating online tools that take greater account of the user's context and use this to help support them in performing the task. This perspective is supported by the *humanistic research strategy* outlined by Oulasvirta (2004), which attempts to find "meaningful uses for context aware technologies", influenced by approaches such as *contextual design*, with its "objective of understanding users in their natural use contexts" (pp. 248). The author sees human needs as providing a sufficient basis for defining a research agenda and consequently argues against use cases that are driven by technology or predictions about how it might be used in the future, but instead suggests that

use scenarios and design must take into account a "holistic understanding of people and their activities", and "of society, users and use situations" (pp. 247).

The role of context in internet usage

Being online can widen the user's potential context greatly. For example, a user can send an email to any other email user if they know the person's address; a benefit of the use of common underlying protocols. This aspect of the internet means current tools tend to make an expanded context possible but at the expense of more immediate context, which is often lost altogether. Some technical approaches to this problem have been developed, and these are discussed below.

A snapshot of context support in current tools

Location and topic as context in web search

Traditionally conventional search engines have treated all users exactly the same (Lawrence, 2000), (Almeida and Almeida, 2004) irrespective of who they are and whether they use the service regularly or are complete newcomers. Whilst this can be seen as beneficial in terms of privacy, it does not allow the user to provide cues about their context through maintaining lists of previous search terms to indicate topics of interest, for example. As search engines continue to develop, these kinds of features are being implemented in services such as Yahoo[2], Ask Jeeves, and A9[3], however, their use is not currently widespread.

Perhaps the only contextual cues that can be reliably given to most search engines involve restricting the search to a specific geographical area, for example by instructing the search engine to only show results from the United Kingdom. However, limiting the scope in this way may merely serve to return pages hosted in the UK, or with a UK specific top-level domain, rather than pages about the UK. Traditionally the web has had no standardised way to declare that a page or object on the web refers to a specific place or location, although developments in the Semantic Web (Berners-Lee et al, 2001) should enable these kinds of statements to be made in a way that is meaningful to both humans and machines.

Gravano, Hatzivassiloglou, and Lichtenstein (2003) draw a distinction between web pages that are of purely local relevance and those of interest to a potentially global audience. They describe an approach that uses machine learning techniques to categorise previously unseen search engine queries on this basis, even though the type of page the user is trying to locate is only expressed implicitly. However, whilst Gravano et al (2003) do mention the user's geographical locality as being of relevance to a search they may perform, this information is not taken into account in their study.

Previous research has also attempted to improve support for context in web search by addressing the topical context of the search term. For example, Lawrence (2000) makes a convincing case for the use of context in general to improve web search, and suggests a greater number of specialised domain-specific search engines may address this problem. Leake and Scherle (2001) go on to propose a system to help users select an appropriate engine from the large numbers available. However, such an architecture could potentially require an infinite number of search engines to cover all domains, suggesting that this solution may not scale well.

In these cases the context relates to the broad topical domain of the search being carried out, rather than the user. Whilst this approach may yield some benefits it is limited in being applicable only to one type of task, in this case locating information using a search engine. Therefore, if a user has an interest in travel, for example, and consequently uses a travel-

[2] http://www.yahoo.com
[3] http://www.a9.com

oriented search engine to locate information in this domain, this contextual cue cannot easily be reused by tools oriented to performing other tasks such as arranging travel. Applying this principle across all tasks users perform online would lead to an explosion of task-specific services or applications in every domain. This would suggest that making use of knowledge related to the user's context, rather than specific to a particular task or domain, would actually be of most value in assisting online task performance.

File and desktop attributes as context

Chirita, Gavriloaie, Ghita, Nejdl, and Paiu (2005) suggest three types of context available on the desktop which may be more fully exploited: email context, such as the discussions that led to the sending of a particular attachment, file hierarchy context, such as the names and structures of folders on a hard disk, and web cache context, such as the files which were downloaded during the same browsing session. Whilst these aspects of context could enable some interesting applications, such as keeping track of the sources of particular objects or knowledge, a number of limitations exist.

The context factors themselves are all application- or protocol-centric, or based on artefacts of current desktop and internet systems. For this reason it is unclear how they may be applicable beyond this specific setting. Perhaps a more scalable approach would be to investigate how contextual factors manifest themselves with regards to each form of digital object (document, image, messages etc.), and how this may be exploited. For example, social context may manifest itself as the sender of a message object, or as a person depicted in a digital photo. This may allow findings to have a broader relevance outside specific aspects of desktop computing systems.

Computing resources as context

Khedr and Karmouch (2004) report on a context ontology and context management agent, and discuss its application in customising a conference environment. However, their definition of context places equivalent emphasis on computational elements ("such as applications running and available services") as it does on aspects of the user ("such as privacy and presence") (pp. 21), a decision which potentially skews the contextual data collected and used by the system.

Scope of knowledge as context

Guha, McCool, and Fikes (2004) outline the history of the term context as used in the field of AI and Knowledge Representation. In this setting it tends to refer to the setting or situation in which a particular piece of knowledge is valid. The authors draw on these ideas in the development of a similar context mechanism for the semantic web, with the aim of avoiding problems in data integration operations stemming from inconsistent uses of terms across providers of semantic web data. Whilst this type of context may be critical in applications that support user tasks online, it is distinct from the user context discussed in this review.

Summary

Clearly several attempts have been made to bring greater context to internet tools. However, they tend to be rather piecemeal, limited in their scope, and in their ability to scale beyond a specific application. Discussions of how context aware applications may be developed that move beyond this piecemeal approach would benefit from reference to more general definitions of context, reviewed below.

Definitions and conceptualisations of context

Review of literature in the areas of human-computer interaction and ubiquitous computing reveal many different ways of defining and conceptualising context, which are discussed here.

Dey, Abowd, and Salber (2001) draw a distinction between the context of the user and the context of the application in a human-computer interaction, which usefully highlights the varied levels on which context can be viewed or the various assumptions on which it can be based. They go on to make a strong argument against attempts to define context *by example* on the basis that these can be hard to interpret and apply, making it hard to decide if a new factor constitutes context or not. However, after dismissing other definitions of context that do so by example, and giving their own reasonable definition of context that takes in the user, the application, and the interaction, they then add that "context is typically the location, identity, and state of people, groups, and computational and physical objects". In doing so they limit the usefulness of their definition and leave themselves open to the same context by example criticisms they made of others.

Dey (2001) provides a definition of context that avoids viewing it as location-centric, but is rather multi-layered: "context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" (pp. 5). Whilst, the assumption of the existence of an application is understandable from an HCI perspective, it reduces the ability of the definition to generalise to other situations. It could be said to reduce its ecological validity, in that real world task performance doesn't necessarily involve an *application*, although it may be influenced by the context of the user as they perform the task.

The corresponding definition of context-aware applications assumes that context is used to provide "relevant information and/or services to the user" (pp. 5), with relevance being defined according to the user's task. Whilst the explicit recognition given to a task-focused view of user interaction is encouraging, the definition implies some degree of passivity on the part of the user (suggesting an emphasis on *push* rather than *pull* services), which could also be seen to limit its scope.

Therefore, despite the attempts of the above authors to move away from narrow conceptualisations of context rooted in factors such as location, the resulting definitions remain overly prescriptive and do not take the user as the primary focus of the context. A more user-centric view of context is crucial if user task performance online is to be supported in a more ecologically valid way, suggesting that a broader view is required.

Fischer (2001) rejects the sensor-based approach to context, and argues that other factors of context need to be taken into consideration: "although the current focus of the ubiquitous computing community is mainly on understanding and handling context that can be sensed automatically in the physical environment, there are numerous other dimensions for context that can and should be taken into account" (pp. 244). Similarly Grudin (2001) provides strong arguments against the definition of Dey et al and the emphasis on location and identity in conceptualisations of context, stating that "this bias is pervasive" (pp. 272).

Dourish (2001) also argues for the need to move beyond seeing context-based computing in traditional ubiquitous computing terms, and refers to social analyses of interaction looking "beyond simply the interaction between an individual user and a computer system" to "the context in which that interaction emerges" (pp. 233). Though not explicitly identified by Dourish, one possible context in which an interaction could emerge is a user performing a task, supporting the idea of contextualised task performance.

Finally, Svanaes (2001) introduces a phenomenological perspective to context-aware technology that "enables a systematic exploration of the design space for context-aware systems - as they appear to the user". Consequently it is argued "that context can never be a property of the world, but that context rather is the horizon within which the user makes sense of the world" (pp. 380). In doing so, Svanaes strongly validates a user-centric approach to defining and using context.

User context factors

Building on the gaps identified in the review above, and in response to the need for greater user focus in defining and using context, Heath, Motta, and Dzbor (2005) propose six user context factors that may be "pertinent when performing tasks online, as they would be in the offline world" (pp. 3). In summary these are: a user's *social context*, services or third parties they *trust*, their history of *previous experiences*, the *resources* they have available for performing the task, *preferences* they hold, or their *current location*. These user context factors are reproduced and described in more detail in Part II of this report. However, for the purpose of this discussion it is worth reviewing how they relate to the definitions cited above.

Crucially these user context factors do not claim to be exhaustive in coverage or to define context, merely to capture aspects of the user's context that may assist in the performance of a task in the offline world, or may serve to shape a task in some way (as the *social context* of guests shapes the *arranging* of a dinner party). With these clarifications in mind it can be seen that the user context factors are not subject to the *context by example* criticism of Dey et al (2001). Furthermore they are largely compatible with the initial part of Dey's (2001) definition of context as "any information that can be used to characterise the situation of an entity" (pp. 5). However, despite being referred to as *user* context factors (thereby implying a human-computer interaction) they are not predicated on the existence of an application, due to the lack of clear *applications* in the offline world and the need to maintain ecological validity.

Whilst this issue is discussed in more detail later, it is worth noting that the user context factors do not exist in isolation from each other. Interactions between them can be observed in that, for example, previous experiences of using a service may influence how much that service is trusted to assist in performance of future tasks.

In considering how these user context factors can support task performance online, one clear advantage of this approach is the potential for enabling reusability of context across tasks. By focusing on use of context as a strategy used in the offline world, and by defining it in terms of the user, then these contextual factors have relevance and validity across different tasks. This avoids the potential situation where an application achieves some degree of context-awareness but only in a very limited setting and in a non-transferable way. Central to this advantage is the ability to describe contextual information related to the factors in a standard, reusable form for use by the full range of online applications.

Existing attempts to provide support for these user context factors in online tools are discussed below. Whilst these primarily take the form of web tools or services, we follow the rationale of Heath, Dzbor et al (2005) that all internet platforms should be considered together, not just the web. The greater focus on social context reflects the future direction of this research. Reasons for this choice are discussed below.

Existing technical support for user context factors

The applications discussed below all attempt to make use of one or more of the user context factors to support task performance online. This section will review and discuss how they do so.

Location

As highlighted by the discussion above, context aware applications often rely on use of location information. Whilst this narrow focus has been widely criticised, location remains an important aspect of a user's context hence its inclusion in Heath, Motta et al's (2005) user context factors. It is worth noting the distinction between the approaches here and the uses of location discussed above under *Location and topic as context in search engine queries*, in which location refers specifically to location-related aspects of the query (such as where to return results from) rather than the location of the user.

Hansen, Bouvin, Christensen et al (2004) describe HyCon, “a framework for context aware hypermedia systems” that uses XLink to represent guided tours, annotations provided by users, and links to other resources. Whilst the system can use awareness of user location to inform web searches for relevant content, thereby avoiding the problems of other tour guide systems where the content comes from a limited pool that must be provided and maintained by a central authority, it is not clear how this representation of location context can be reused in support of tasks other than locating information and the serendipitous discovery characteristic of the grazing task.

Duri, Cole, Munson, and Christensen (2001) raise the issue that the end user experience of location aware applications will increase in importance as more mobile commerce services become available; a view which is compatible with the user-centred perspective adopted in this report. To help address these issues and give users straightforward access to location based services that change according to their location, Duri et al propose *dynamic bookmarks* which they define as “descriptions of services, which are bound to actual, registered, services as a user's location changes” and *location domains* which “provide meaningful location context for location-aware services” (pp. 20). These mechanisms would seem to provide a viable way for users to express their needs, whilst the responsibility for finding relevant services that meet those needs falls to the application.

In conclusion, whilst some promising approaches exist, it would be encouraging to see if a generic user location framework could be developed that can scale across tasks whilst remaining driven by the user and within their control.

Preferences

Considerable literature exists regarding personalisation in web search, but largely this relies on developing profiles of users over time based on their search behaviour. One example of this approach is reported by Middleton, De Roure, and Shadbolt (2001) whose system (applied to the domain of academic papers) analyses a user’s browsing behaviour before using “supervised machine-learning techniques coupled with an ontological representation to extract user preferences”.

Whilst these approaches may implicitly reveal aspects of a user’s preferences, these findings cannot always be validated by the user (although Middleton et al do express a desire to incorporate user feedback to improve recommendations), and can rarely be exported in a meaningful way for use in other systems.

Looking specifically at musical tastes, which can be seen as one form of user preference, Celma, Ramirez, and Herrera (2005) describe a system that takes explicit declarations of preference for certain musicians/bands (using the *foaf:interest* property) and uses this as a basis to recommend other artists the user may like. This system provides a seemingly viable model for user expression of preferences that can then be utilised by other services to assist in task performance online. It would be interesting to investigate the application of this approach to other domains beyond music. To do so would require vocabularies for expressing other preferences that user’s may have (perhaps such as dietary requirements), and definitions of how to interpret and reason with these.

Resources

Within the field of Computer Supported Collaborative Working (CSCW), Voss and Kreifelts (1997) describe a system that attempts to take social context, history, trust, and resource factors into consideration. Known as SOaP (an abbreviation of Social Agent Platform), the system seeks to aid users in locating web resources of interest by supplementing regular search results with items that trusted team members have seen before and recommended. The (cognitive) resources required by the user are deemed to be lower due to the use of agents to aggregate and distribute the items.

Theories of *information foraging* Pirolli and Card (1998) suggest that people will make use of strategies that minimise the resources required in locating information, supporting the existence of systems such as SOaP. Whilst this work on information foraging only applies directly to the tasks of locating and exploring, the principle that users will seek strategies that minimise the resources required to perform a task, or that will adapt the strategy used depending on the resources available, is likely to apply to other tasks in the classification presented by Heath, Dzbor et al (2005).

Further research into how user's may express the resources they have available for performing a task may be informed by research into *presence*, which looks at issues such as how people describe their availability and attention levels to others. This area may be able to contribute ideas about how user resources can be characterised.

History

Despite being party to almost all user actions on the web, the average web browser makes little use of this information to assist in the performance of tasks online. One aspect that is widely but superficially taken into consideration is a user's previous experience, with browsers keeping a record of previously visited sites and values entered in form fields. This is generally referred to as the user's *history*. However, despite the majority of pages that a user views having already been seen by them before, current history tools in browsers do not support users well in re-accessing pages this way (Tauscher and Greenberg, 1997). Users often have to recreate the steps they originally performed to reach a particular page, meaning that this history data is frequently wasted. Similarly, records of entries made in form fields are used simply to prompt the user with possible inputs they may wish to make; additional services are not readily available that further exploit this data for the user's benefit. Furthermore, browsers are certainly not able to keep a record of what tasks have been performed in the past and how they were completed. Consequently, as with information about pages viewed, solutions cannot be reused unless the user is able to manually reconstruct all stages of the process.

Looking specifically at data from a browsing session as a source of contextual information, Chakrabarti, Srivastava, Subramanyam et al (2000) describe a *browser assistant* applet that merges bookmarks and history, and attempts to provide topical structure to this data. Users begin by importing existing bookmarks, which may already be topically grouped into folders. These serve as the starting point for the user's personal topic taxonomy. Subsequent page requests are captured by the applet, the page is retrieved, and then analysed according to its similarity (in terms of textual content) to other pages previously requested by the user. On the basis of this analysis, the applet makes recommendations about new ways to cluster and subdivide folders in the topic taxonomy. Whilst this system may assist in accessing previously viewed pages or in understanding the topical structure of pages viewed or bookmarked, it does not appear to provide rich contextual data that is reusable in other applications.

Building on the ontology-driven document enrichment approach described by Motta, Buckingham Shum, and Domingue (2000) is the web browser plug-in *Magpie*, which supports the browsing and interpretation of web pages by highlighting known entities that occur on the page being viewed Domingue, Dzbor, and Motta (2003). The plug-in is complemented by a wider framework of server-side components that provide *on-demand* and *trigger* semantic services, in response to occurrences of named entities in a populated ontology. One such service is a mechanism for semantic review of browse history recorded in a *semantic log KB* Domingue, Dzbor, and Motta (2004). Recognised entities (from the ontology) occurring in a viewed page are sent to the semantic log KB by the browser plug-in, and a record is kept that they have been seen. Whilst the indexing and storage resources required with this approach are relatively low, it does not allow for subsequent re-indexing according to new ontologies, which could limit the applicability of history data to other domains.

Not limiting a user's history just to one web site or even the web as a whole, Dumais, Cutrell, Cadiz et al (2003) describe a system that captures user history across a whole range of desktop applications. Known as *Stuff I've Seen*, the system indexes the contents of emails, web pages, documents, and appointments, amongst others, and attempts to improve the

user's ability to retrieve items based on the terms indexed. Whilst evidence suggests the application is of value to users, it remains a closed system that may be hard to extend.

More effective reuse of users' experiences, such as browsing history and previous solutions to tasks performed online, could properly recognise Bush's (1945) vision of *The Memex*; a means for capturing and managing individual knowledge, and sharing it with others. Whilst the envisioned system uses a spatial metaphor to describe trails through related items, semantics are inherent to the system in the form of associative indexing; "the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another". This original vision has often been cited since the web came into being.

From reviewing the tools discussed above it is clear that there are a number of possible mechanisms for indexing and retrieval in history systems, each with varying degrees of semantics. There are also various levels at which these applications can be implemented; in the web browser, as desktop applications, as *webtop* applications, or as combinations of several. One thing that is not clear from this review is how reusable the captured histories are across different tasks and outside of their host applications. However, it would appear that the types of history information gathered are still relatively narrow and application specific.

Trust

Heath, Motta et al (2005) define this context factor as "trust the user has in third parties; individuals, groups, organisations or services that may be able to assist in performing the task" (pp. 3). In this discussion it is worth distinguishing between trust the user has in other people from the trust they may have in services. Typically services perform just a few specific operations on behalf of the user, such as dispatching books in exchange for payment, providing travel agency services for example. For this reason the user only needs to trust the service to fulfil certain specific actions and behave in certain specific ways, such as deliver goods as agreed, not use payment details fraudulently etc. Contrast this with describing trust in other human beings, who can conceivably perform a far wider range of actions, and the nature of the trust relationship becomes far more complex. One may trust a friend to recommend interesting films to see, but not to borrow large sums of money and return it.

In the offline world, people use varied cues to gauge whether they should trust another individual or organisation; some of these are taken in isolation, such as the appearance of a shop or a person, whilst some are derived from other context factors such as recommendations or previous experiences with the party in question. The abstract nature of the web makes these sorts of assessments more difficult. This can be a problematic issue, especially considering the importance of trusted third parties in contributing to a user's context when performing tasks.

A number of attempts have been made to create trust or reputation systems online that address both trust in people and in services. Some well-known organisations have set up schemes for validating online retailers to demonstrate their trustworthiness to users. These generally manifest themselves through the display of *seals of approval* on approved web sites. However, the criteria used to assess this trustworthiness are often not clear to the end user and may carry less weight than validations obtained from highly trusted sources such as friends or family. Online auction services such as eBay[4] allow users to rate each other based on previous transactions, as a means of indicating trustworthiness. Leaving aside issues of constructing false identities that appear trustworthy, this system suffers from the same limitations as Amazon[5] recommendations (discussed below), in that the knowledge held by one site cannot be used by others.

Considering how these schemes support the representation and use of user context online, it would appear there are some limitations. Whilst the approved trader schemes may give an indication of which services a user should trust, they do not allow them to express which

[4] http://www.ebay.co.uk
[5] http://www.amazon.co.uk

services they do already trust. Consequently the user must demonstrate their trust by actively choosing which service to use, rather than allowing applications or agents to select automatically from their list of trusted services.

Researchers such as Golbeck and Hendler (2004) have used semantic web technologies to enable the expression of trust relations between individuals online. These approaches may go some way to addressing the limitations outlined above, and are discussed in more detail below under the heading *Modelling trust between individuals*.

Social

The importance of social context in assisting task performance has been raised by a number of authors. For example, a classic study by Granovetter (1973) investigates the role of social networks in the diffusion of information (amongst other things), and finds that they are of great importance. However, the key finding is that it is not simply the immediate nodes in a person's social network that are of use, but also the weaker more distant ties. This finding is attributed to the lesser overlap in friendship circles of people who are only weakly tied. In turn this leads to an increased pool of information being available to weakly tied parties.

Raising the agenda of social processes in the field of computing, Schuler (1994) defines *social computing* as "any type of computing in which software serves as an intermediary or a focus for a social relation". The author identifies a need to emphasise the role of computing in social processes, and a need to address social factors in computing. This view is supported by Erickson and Kellogg (2000) who introduce the concept of *social translucence*, defined as "an approach to designing digital systems that emphasizes making social information visible within the system" (pp. 61). Two dimensions of social translucence are identified: "visibility of social information", and "visibility of social interactions". The authors argue that support for neither of these factors can be taken for granted in computing systems.

Grudin (1994) charts the evolution of single user systems into groupware, attributed mainly to the wider availability of and familiarity with computing resources, and to an increase in connectivity between these resources. The author goes on to argue that social software should be integrated into existing applications, a view that resonates with the social navigation ideas discussed below.

Social navigation

In defining the concept of *social navigation*, Dieberger, Dourish, Hook et al (2000) draw links between *real-world* navigation methods and how navigation should be supported online. The point is made that navigation aids from the offline world such as maps and guides have been used to support navigation online, whereas social navigation principles have been slow to be implemented in computing systems, despite their regular use in the offline world.

Dieberger, Hook, Svensson et al (2001) develop these ideas further and draw a distinction between *global social navigation* where the principles are adopted in all applications, and a more *task-oriented* approach where social navigation is best utilised for specific tasks. The perspective taken in this report could be seen as a hybrid of these two options, in that the integration of context into internet tools is seen as advantageous, but this contextual information is used in support of specific tasks.

Svensson, Hook, Laaksolahti et al (2001) apply the social navigation idea to the development of a recipe recommendation system. This involves the use of the *social navigator* toolkit that allows *flags* to be used to signal aspects of a person's context. However these flags do not have any predefined semantics. In conclusion, whilst the principles of social navigation add weight to the argument for greater use of context in online tools, by definition the field only addresses some of the tasks outlined by Heath, Dzbor et al (2005), such as locating and exploring.

Collaborative filtering recommender systems

One service that has made effective use of community and social factors to help people perform tasks is the online retailer Amazon. As well as keeping a record of a customer's purchase history and using this to recommend additional items they might like to buy, the service uses a collaborative filtering mechanism to suggest items that others who match the customer's shopping profile have bought. Linden, Smith, and York (2003), describe how the recommendations system matches purchased or rated items to other items, with the similarity rating between items being based on their likelihood of having been purchased together. Compared to traditional collaborative filtering methods, this approach remains feasible even with large collections of products and customers.

In terms of supporting task performance, not only can the methods described above be used in locating specific items, it also supports the grazing task identified in Heath, Dzbor et al (2005); users may not have a specific goal or item in mind, merely an expectation that interesting items may be encountered if they use the system. However, despite subjective reports from users of the success of this system, its coverage is limited to items that exist within the retailer's database and that can be bought through their web site. There is no standardised means to expose one's purchase history (if desired) for other services to make use of, and no way to define specific social context the customer would like to make use of for providing recommendations, so the context is restricted and specific to that one system.

Shardanand and Maes (1995) deploy similar socially oriented filtering techniques to power a music recommendation system called *Ringo*. The authors compare four different algorithms to find the most effective. However, despite reporting positive results, similar criticisms apply to this system as to Amazon recommendations. *Ringo* does not allow the user to define their social context and use this for generation of the recommendations. Instead the user receives suggestions based on the ratings of others who fit the same profile, and who may or may not be known to them. There is not scope for selecting groups of people to draw recommendations from based on other criteria.

So, whilst collaborative filtering recommender systems can produce some useful results, they are limited by the closed world systems in which they are implemented. These closed worlds limit the ability of the user to define their context and how it is used in task performance, they limit the ability to reuse contextual data that may have been created by the user, and they restrict the domains in which the contextual data can be applied.

Conclusion

The technical approaches reviewed above represent a wide range of attempts to address the user context factors discussed in this report. However, in conclusion, it would appear that there remains a lack of adequate support for representing and making use of these context factors in online task performance. To increase the levels of support it may be necessary to model each of the context factors in more detail. The following section existing attempts to represent and model social context online.

Modelling social context on the semantic web

Central to any discussion of social networks and the semantic web is the RDF vocabulary *Friend of a Friend* (FOAF) (Brickley and Miller, 2004). Ding, Finin, and Joshi (2005) report that there are nearly one million instances of the *foaf:Person* class on the web, distributed among roughly 45,000 documents. The following two sections will examine how FOAF and other methods have been used to represent aspects of social networks.

Modelling the nature of social relationships

The large amounts of FOAF files available on the web have been used as a source of data for performing Social Network Analysis. For example, Mika (2004) uses FOAF files plus

information mined from the conventional web to analyse the social networks that make up the semantic web research community.

However, different perspectives exist about how best to represent people's social networks and the relationships between people on the semantic web. The basic unit of defining relationships in FOAF is the *foaf:knows* relation, simply used to state that Person A *knows* Person B. Whilst the simplicity of the description is undoubtedly an advantage, there may also be limitations due to a lack of expressivity.

Other vocabularies have been proposed that go beyond the shallow semantics of foaf:knows, and describe more detailed relationships between people, such as the *Relationship* vocabulary[6]. However, this appears to have severe limitations. For example, the vocabulary can never be descriptive enough to fully capture how two people know each other. In their discussion of various perspectives on the nature of the semantic web, Marshall and Shipman (2003) note that it can prove hard for computers to exchange information about concepts which are oriented towards humans due to issues of representation, and this certainly appears to be the case with describing the nature of relationships between people.

Furthermore, it isn't immediately clear how a relationship defined very precisely could actually be used in an application or to support a user task. It raises the following questions: What inferences can be made based on a particular relation? Can more reasoning be performed if the full nature of the relationship is actually known? Is knowing the nature of the relationship at all informative as the basis for helping with task performance? To resolve this problem requires a greater understanding of how people actually use knowledge of relations in a social network when performing tasks. This would enable logical rules to be developed that defined the possible actions stemming from a particular relationship. Also it may transpire that relationships are better discussed in functional terms: "*what does this link in the network actually allow me to do?"*

It is worth noting that describing family relationships may prove more feasible, as they are often more neatly defined than non-family social relations. However, it may be that these sorts of descriptions are culturally dependent, so assumptions of equivalence across cultures should not be made. Semantic descriptions of family relations may have clear applications in areas such as genealogical research, however it remains unclear what how they could be used in supporting more generic tasks.

Modelling trust between individuals

Building on the widespread availability of FOAF files, Golbeck and Hendler (2004) describe a vocabulary that allows users to rate the reputation of trustworthiness of other people in the network. These ratings are then used as the basis for an email sorting and filtering application. Whilst the simplicity of the system can be seen as beneficial, authors such as Kalfoglou et al (2004) have questioned the value of such trust ratings on the basis that they are de-contextualised. It is not possible to convey the setting or situation in which this rating applies, and these can vary widely, as the example given earlier regarding films and money lending suggests.

Gray, Seigneur, Chen et al (2003) use small world theories as the basis for trust systems that can scale to massively distributed networked applications, of the kind that will increase as mobile technology becomes ever more widespread. The approach is based on consideration of how trust decisions are made in the offline world, taking in concepts such as risk and recognition. A key distinction to make is that Gray et al go on to focus on entities such as users can deploy locally available information to determine the shortest path through a network for propagation of information.

[6] http://vocab.org/relationship/

In work described by Gil and Ratnakar (2002), users create annotations about web resources, which are then used to determine the degree of trust that exists in the content of the document. The authors express an interest in extending the system to descriptions of the reliability and trustworthiness of services, thereby supporting the distinction raised earlier between trust in other people and trust in services, and providing a solution that addresses both.

Richardson, Agrawal, and Domingos (2003) describe work based on the premise that "a user's belief in a statement should be a function of her trust in the sources providing it". They go on to emphasise that trust is subjective in nature and therefore cannot be assigned to an entity by a central source, and that furthermore, no one entity will know the trustworthiness of every source. On this basis, the authors propose a distributed web of trust approach, where each user assigns trust values to a small number of other users. These are then aggregated into trust values, not globally as a measure of each individual by the whole group, but as a personalised measure of their own trust in every other member. Because these are specific to each user, they may vary widely, embodying a user-centred model of trust that is compatible with the perspective taken in this report. Richardson, Agrawal, and Domingos (2003) also plan to enable "users to specify topic-specific trusts", which may go some way to addressing the criticism of Kalfoglou et al (2004) regarding the trust vocabulary.

Conclusions and gap analysis

The review and discussion above highlights a shortage of tools adapted to supporting user tasks online, and making use of the context factors that can help in their performance. In particular it would appear that greater attention could be paid to offline task performance strategies, such as use of context, to inform the development of tools. In contrast with other conceptualisations of context, user context is poorly represented, or the definitions do not take into account a full range of contextual factors that might be relevant to the user.

Looking specifically at user social context, and a community knowledge-centric approach to task-performance, demonstrates a gap in the research regarding modelling of social networks and how this might be deployed online. Certain questions are raised, such as: is it possible to map more detailed social relationships? In what ways can more detailed relationship definitions be used? Does knowing more about the nature of a relationship actually bring benefits, value, and utility in supporting performance of tasks? How can the user assess who will be useful in helping them perform a particular task in a particular domain? What is the role of inferred relationships that haven't been validated by a human? To what extent can these be trusted and used?

These issues demonstrate a need for greater understanding of how people use their social networks to help them perform tasks, and how these factors may be modelled. This increased understanding should then be complemented by an infrastructure for community knowledge modelling, capture, and use, and tools for tasks such as knowledge sharing within social networks.

Limitations of this review

Further reading is required to fully understand previous findings on the functioning of social networks. In particular, review of additional research into how trust and reputation are mediated by social networks is required, and will provide a useful foundation for the user study and attempts to model these, as described in the research proposal.

In depth review of work carried out related to certain context factors (such as location and preferences) has deliberately not been included in this report. Whilst this would be a valid and interesting area for future research it is beyond the scope of this research at this time.

Part II. Specific practical outputs to date

Introduction

Practical work to date has consisted of several strands of activity. The first of these has been the *creation of conceptual artefacts* in the form of classifications of user tasks and context factors detailed below. The second has been *planning of a social context application*, as detailed in the research proposal in Part III of this report. This has included production of a motivating scenario and system walkthrough, plus identification of the system's architectural components, inputs, and outputs. These can be found integrated into the research proposal under the heading of *Research component 2*.

The third strand of practical work has focused on *identifying and assessing relevant tools and technologies* (such as ontologies, vocabularies, libraries, and frameworks) that may be used in the development of the social context application. Discussion of how these may be utilised is mainly embedded in the research proposal under the heading of *Methods of implementation and technologies used*, within *Research component 2*, whilst the detailed analysis of existing travel ontologies is presented later in this section. The fourth strand of activity has involved *acquiring new technical skills* to support the application development described in the research proposal. Development of these skills is discussed at the end of this section.

A user-oriented classification of online tasks

Drawing on the work described in the literature review, and on the discussion of the limitations of previous ways of conceptualising online tasks, the following classification is proposed as a model of tasks users perform online. This classification may allow tools to be developed that support user tasks online more directly.

Task	Definition	Example
Locating	Looking for an object or chunk of information which is known or expected to exist; it may or may not have been seen before by the user.	Locating an article from a journal, an image for a school project, or information about a book a friend recommended.
Exploring	Gathering information about a specific concept or entity to gain understanding or background knowledge of that concept or entity.	Exploring a philosophical theory to understand its central tenets; getting background information about an organisation before a job interview.
Monitoring	Checking known sources that are expected to change, with the express intention of detecting the occurrence and nature of changes.	Monitoring news web sites during an election; checking email accounts for new messages; watching discussion fora for new ideas or information.
Grazing	Moving speculatively between sources with no specific goal in mind, but an expectation that items of interest may be encountered.	Following links that spark your interest on someone's web log, just to see what you find.
Sharing	Making an object or chunk of information available to others.	Sharing holiday photos through an online photo album; uploading a journal article to your personal web site.
Notifying	Informing others of an event in time or a change of state.	Emailing a group of friends to tell them you will be going to a concert at the

		weekend.
Asserting	Making statements of fact or opinion.	Writing on your web site that you like a certain film or artist, or that you own a certain book.
Discussing	Exchanging knowledge and opinions with others, on a specific topic.	Posting a comment on a discussion forum stating that you disagree with a previous post, explaining why, and then receiving responses from others.
Evaluating	Determining whether a particular piece of information is true, or assessing a number of alternative options.	Choosing which film to see at the weekend, based on what's showing, where, and at what time.
Arranging	Coordinating with third parties to ensure that something will take place or will be possible at a certain time.	Arranging travel and accommodation for an international conference.
Transacting	Transferring money or credit between two locations; may or may not have some consequence in the offline world.	Paying a bill, or transferring money between accounts.

Table 1: a user-oriented classification of online tasks

Assumptions made in the classification

The classification presented above is based on a number of assumptions. These assumptions may reflect progress beyond the limitations of previous classifications, or assumptions made about ways in which internet usage will develop.

Clearly the classification presented here addresses a wider range of user tasks than those described previously. One reason for this greater coverage is that it doesn't make assumptions about the task being performed using a particular internet platform (such as the web), only that the user is online by way of some form of internet connected device. For example, notifying might take place via email, and discussing could take the form of threads on a web discussion board or an instant messaging conversation, but ultimately the protocol being used should be transparent to the user. This serves to not limit the classification to a specific domain such as searching or a specific application such as a conventional web search engine or web browser.

The connectivity available to the user's internet connected device is assumed to be permanent, and (for all practical purposes) not restricted in terms of bandwidth. Similarly the user is assumed to have access to virtually unlimited storage which is permanently networked, abstract with regard to host machine or location (in much the same way as logical volumes work in an operating system), and crucially, is addressable using core internet and web principles such as URIs. This does not assume that a user's storage is publicly accessible, only that common methods can be used for working both with user files and with objects on the public internet. Redundancy of online services is assumed, such that objects and services can be thought of as always available. These assumptions account for the absence from this classification of tasks such as *Save to disk* that have been identified by previous authors Byrne et al (1999). If the user has permanent connectivity and universally addressable storage that is not tied to a particular machine, the notion of saving a copy for *offline* access becomes obsolete.

In a similar vein, the task of *locating* does not make any assumptions about whether an object has been seen previously by a user or not; in this sense it could be seen as combining

aspects of finding new objects and retrieving those that have been previously seen. Terms such as retrieving are avoided as they imply the use of local archives or storage.

Finally, the most significant assumption is that the tasks presented here are seen to bear relations to people's actions in the offline world. In that sense they strive to echo processes or behaviours that are familiar from variety of use contexts, rather than specific to one environment such as the internet. In doing so they could be seen to possess greater ecological validity than tasks identified in previous research.

Linked tasks

During any one online session a user may perform a number of tasks that, whilst distinct, are in some way related; these could be thought of as *linked tasks*. For example, a user may have heard that a concert is on in the city where she lives. She would like to go to the concert, and so uses a listings web site to find out that it starts at 8pm. Thinking that her friends might like to go as well, she then emails them to let them know about the concert, mentioning the start time. In this case the first task is clearly an example of *locating*, as you set out to find a certain piece of information, whilst the second task constitutes *notifying*. Here the two tasks bear a thematic relationship but remain tasks in their own right, each addressing a particular goal. Similarly, monitoring a news web site may reveal a story of interest that results in the user grazing related sites with the expectation of finding other relevant items. Shopping online can be seen as a further example of linked tasks. The act of paying for goods or services can be classified as transacting, and this may be preceded by locating a specific item to purchase or evaluating a number of different options.

User context factors in task performance online

A number of aspects of a user's context can be identified that may have significant roles to play in shaping the nature of the task and the way in which it's performed. These factors are detailed and defined below.

Context factor	Definition
Social	groups of individuals a user is part of or identifies with, such as friends, family, colleagues, or others with shared interests.
Trust	trust the user has in third parties; individuals, groups, organisations or services that may be able to assist in performing the task.
History	actions the user has taken in the past; solutions to previously performed tasks; experiences of these tasks and solutions.
Resources	resources such as time, money, attention or tools that may be required in performing the task but may vary in their availability
Preferences	values or opinions that a user holds that might effect how a task is performed, or what solutions might be acceptable
Location	the user's physical or geographical location.

Table 2: a conceptualisation of user context factors when completing tasks online

These factors can all be seen as properties of the user, with varying degrees of stability and persistence over time. For example, a person's previous experiences do not change over time, though they may be added to; the social networks a person associates with are likely to depend on the situation, but the links in the network are likely to change at a fairly slow rate; in contrast, the resources a person has available for performing tasks are likely to vary frequently.

Crucially these context factors are likely to manifest themselves differently depending on the task being performed. For example, in tasks such as notifying or sharing, members of a user's social network may be seen as the audience for the task or the beneficiaries of its outcome, rather than sources of assistance; discussing on the other hand might involve contribution from all individuals, presumably for mutual benefit. Taking the factor of trust as an example, the extent to which a user trusts a third party web site may be of great significance if they are carrying out a transacting task such as paying for goods or transferring money. However, in contrast, if they are exploring a controversial topic and simply want to survey a broad range of opinions it may not matter whether they trust the sources they find or not.

Interactions between factors

Consideration of the factors above demonstrates that they do not exist in isolation of each other. For example, as mentioned in the analysis of a semantic web scenario below, a relationship can be seen between social context and trust, in that a recommendation from a friend is likely to either increase or decrease trust in the recommended entity, depending on how the friend's recommendations are perceived. Similarly, previous experience with a person or a service is likely to affect how trusted they are in future interactions. Taking location as an example, then the user's current location may impact on their resources available for performing a task, if they are in crowded street rather than in a quiet office, for example.

Analysis of user context factors in a semantic web scenario

A widely cited scenario used to illustrate the potential of the semantic web is given in Berners-Lee et al (2001) and concerns two siblings (Pete and Lucy) arranging healthcare treatment for their mother, with the assistance of semantic web agents. Analysis of the scenario highlights the crucial role that user context factors play in defining and supporting the performance of the task. Whilst the solutions used in the scenario are semantic web solutions, the tasks facing Pete and Lucy are equally relevant to the conventional web. A key factor to note is that whilst the user context factors can be observed in the scenario, they are not all well represented, if at all.

From the outset, the task is defined by the *social context* of Pete and Lucy in that it concerns their mother and they need to perform the task together or in consultation with each other. The fact that “Lucy’s agent has complete trust in Pete’s agent in the context of the present task” (pp. 36) suggests an interaction between social factors and *trust*.

History of previous experiences or solutions to tasks is shown to be of value when Pete instructs his agent to redo the search carried out by Lucy’s agent, and Lucy’s agent supplies “shortcuts to the data it had already sorted through” (pp. 36). In this case the history being reused belongs to someone else (or their agent), however if Pete’s agent needed to perform the task again in the future it could now reuse its own solution.

In addition to the trust relationship between the siblings’ agents mentioned above, the scenario also refers to *trusted services* that give ratings of healthcare providers. Here Lucy (through her agent) trusts a third party service to supply accurate information that can be used in performing the task. If a rating service was not available that Pete and Lucy trusted, the quality of providers would have to be verified by other means, complicating the task considerably.

Many examples exist in the scenario of the importance of *preferences* as a context factor. The first of these relates directly to the ratings of healthcare providers, where Lucy’s agent only retrieves details of providers rated excellent or very good; presumably their preferred standard for providers that may treat their mother. Pete’s instruction that his agent redo the search carried out by Lucy’s agent results from the proposed solution not meeting his preferences; he doesn’t want to drive across town and return at rush hour, so the search is redone with stricter preferences, slightly redefining the task in the process. This also serves to highlight the significance of *location* as a context factor, with distance of the hospital from their mother’s home and from Pete’s office both needing to be taken into consideration.

At different stages of the scenario Pete clearly has differing levels of *resources* available to dedicate to the task. When his agent first returns an alternative solution to the task, he has little attention for verifying every aspect of how the task was performed, and mutters “spare me the details” (pp. 36). Instead he defers this until later when he has the time to review the record of how his agent reached the solution.

Assessment of relevant tools and technologies

A wide range of tools and technologies (such as ontologies, vocabularies, libraries, and frameworks) have been assessed to determine their potential utility in the application development aspect of this research (described in the research proposal in Part III of this report). To ensure they are considered in context, the conclusions of these assessments are primarily embedded within sections of the research proposal under *Methods of implementation and technologies used*. However, comparison of existing travel vocabularies and ontologies is presented here in detail and summarised in the narrative of the research proposal.

Comparison of travel vocabularies and ontologies

Name	Hotels	Food	Places	Sights	Notes
DAML Geofile			X		Potentially useful breakdown of locations at varied granularities, but has military not traveller focus
DAML Country Codes/CIA World Factbook			X		Useful at a high level, but no granularity beyond the country level.
Basic Geo Vocabulary			X		De facto standard for describing longitude and latitude, but no finer detail. Useful for giving location of other objects
Reuters Regions Vocabulary			X		Potentially useful due to worldwide coverage, but has political not traveller focus
DAML Airport Ontology			X		Potentially useful in a wider travel ontology, but otherwise too specific
DAML Travel Itinerary Ontology	X	X	X		Broad coverage but granularity at a level appropriate to the travel industry, not to travellers sharing recommendations
TAGA Travel Ontology	X				Oriented to reservations within the travel industry. Hotel coverage is shallow
Tsinghua Travel Ontology	X	X	X		Potentially useful. Broad but shallow. Would need to be supplemented
ChefMoz RDF Vocabulary		X			Detailed description of restaurants possible. Very useful, but mixes restaurant info with reviews.
OpenGuides RDF Descriptions		X	X	X	Fairly rich descriptions of various objects, but mainly using other namespaces. Very useful as a data source.

Table 3. Comparison of existing travel vocabularies and ontologies, and their ability to describe "travel objects"

Development of technical skills

To support development of the social context application proposed in Part III, a number of technical skills have been developed or built upon. Firstly, considerable effort has been put into becoming familiar with the Resource Description Framework (RDF), mainly achieved through using RDF vocabularies to describe people, places, photographs, and resources. This has been complemented by investigating tools that can be used to discover, store, manipulate, and produce RDF data, some of which are discussed in the research proposal, under the *Methods of implementation and technologies used* heading.

A number of smaller projects undertaken in KMi have provided the opportunity to develop new skills in a number of other technologies highly relevant to this research. For example, development of an application to retrieve cinema listings from web sites through a Jabber interface has provided experience in using Perl[7] to scrape web sites and parse the results. This approach could be utilised to create semantic descriptions of existing web resources as may be required for population of the domain knowledge base, as detailed in the research proposal.

The *Bookshelf* project undertaken in KMi uses PHP[8] to query the Amazon Web Services API[9] via the REST approach, retrieving XML records which are then cached, processed, and transformed using XSLT for display to the user. Whilst building on existing knowledge and experience, skills in these technologies were all developed specifically for this project, in a short period of time.

In addition, the Bookshelf project provides a valuable test data set for use in this research. Whilst the travel-specific components of the application described in the research proposal are being developed, the bookshelf project could provide a dataset with which to test generic aspects of the system related to sharing knowledge within a social network. This would allow valuable lessons to be learned that could inform the development of the main applications.

[7] http://www.perl.org
[8] http://www.php.net
[9] http://webservices.amazon.com

Part III. Research Proposal

Statement of research question

“Can a person’s social context be modelled, captured, and utilised to facilitate locating and exploring information online?”

Definition of terms

Social context in this case is defined as the social networks that the individual is part of or identifies with, and the people that make up these networks. The relation of the individual to society as a whole is not considered part of the social context in this research. In recognition of the fact that social networks are primarily a construction of the individual, no assumptions are made or criteria set at this stage about the nature of the relationships between the individual and members of their social networks, or about the origin of the relationship. Consequently the social context may encompass family members, friends, colleagues, and other acquaintances identified. Furthermore, no assumptions are made at this stage about how particular classes of network members may contribute to task performance, if indeed any distinction is made by the user on the basis of relationship type. These decisions will be informed by the pilot study described below.

The terms *locating* and *exploring* refer to the tasks defined by Heath, Dzbor et al (2005) respectively as “looking for an object or chunk of information which is known or expected to exist; it may or may not have been seen before by the user” and “gathering information about a specific concept or entity to gain understanding or background knowledge of that concept or entity” (pp. 3). These definitions are adopted without modification for use in this research.

Justification for research question

The question presented above is important and worth asking for the following reasons. It addresses a clear *user need* that is applicable to all internet users, specifically how to avoid information overload in an era where information is widely available. The research has the potential to *lower barriers* between the online and offline worlds, by allowing processes and principles developed offline over many years to be more easily used in an online environment. Furthermore the outcomes of the research have *wider applicability* in that they can inform how other user context factors may also be used to support user tasks online.

Considering these issues in more depth, the importance of developing technical systems that are driven by a clear user need is highlighted by Oulasvirta (2004). As increasing amounts of information are made available online the concept of information overload has come to the fore, and overcoming this problem presents a clear user need. Information overload may not occur to the same degree in the offline world, but literature shows that when locating information people make use of their social networks (e.g. Granovetter, 1973). This research seeks to make these *word of mouth* principles usable online, providing a more contextualised online experience for the user, which the literature review demonstrates is not currently the case. Improving support for these principles fulfils the need to integrate social aspects into all forms of computing applications (identified by Grudin, 1994), and is compatible with making tools that are more translucent to social processes, as argued by Erickson and Kellogg (2000).

It is not currently apparent whether people use their social network to obtain information specifically from others who are known and trusted, or simply because it serves as a convenient filter when large amounts of information are available. Investigating this research

question has the potential to further knowledge about how people use their social networks to help them perform tasks. Investigations into how this contextual information can they be modelled, captured and utilised online will provide a useful test bed for the use of semantic web technologies in supporting contextually aware applications, an area which has not been extensively investigated to date. Successful outcomes of the research could enable a new generation of socially contextually aware tools, the principles of which could scale to other context factors and enable new ways of approaching contextually aware tools and services.

The social context factor has been chosen for this research on the following basis. Firstly there is a strong and growing research interest in the area, as evidence by the publication of articles concerning analysis and use of social networks online, and the presence of this topic in workshop and conference programmes. Secondly from a technical perspective there is significant infrastructure in place already to support the use of social network information online, such as the large amount of data on the semantic web that uses the Friend of a Friend vocabulary. However, this research will attempt to illuminate how relationships may best be described and utilised for practical applications. Thirdly, anecdotal evidence suggests that there is strong user interest in expressing and using social network information, but a lack of practical applications with which to do so and with which to motivate others to participate.

The tasks of locating and exploring have been chosen from among the eleven identified by Heath, Dzbor et al (2005) for a number of reasons. From a user perspective, they constitute a modality that users are already familiar with in the online world (through use of search engines for example), although the average user may refer to them in different terms. They also serve as a feasible test bed that builds on considerable existing experience and literature in the field, generally developed using conventional web and information retrieval techniques. Socially contextually aware applications could have a noticeable impact in comparison to conventional tools, which provide a useful benchmark for comparison. If social context can be utilised to support these tasks then its use could be extended to other tasks.

Conceptual representation of the research area

Figure 1 below unifies the user tasks online and the user context factors, illustrated with reference to the use of social context to support the locating and exploring tasks.

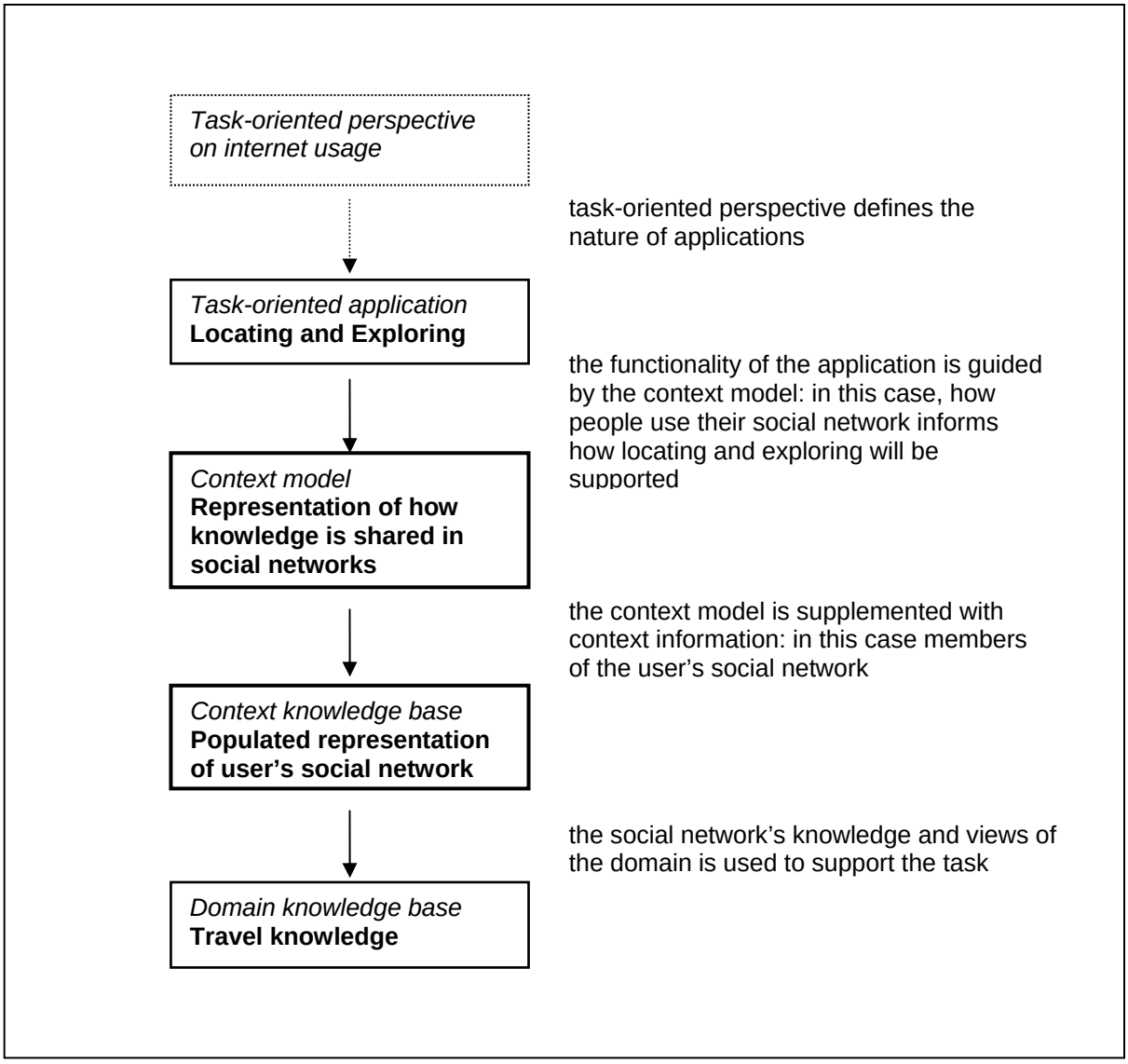

Figure 1: Conceptual representation of the research area *contextualised task performance online* with illustrative examples

Motivating scenario

The research proposed here can be captured from a user point of view by the following scenario that, in combination with the gaps identified by the literature review, motivates the proposed research.

Locating and exploring travel knowledge via a social network

Steve is planning a backpacking trip to New York and Boston. He knows that a number of friends, family and colleagues have been to the USA before, so he wants to use their knowledge and experience to help plan his trip. However, they're busy people, and so is Steve; he doesn't want to have to construct an email asking each of them for all their recommendations, and he knows some of them have already logged their trips in detail on different travel sites, he just doesn't know where to find these reviews.

Fortunately there's a solution; some of the travel sites provide users' reviews in a standard machine readable format. This format includes information about the subject of the review

(perhaps a hotel, a restaurant, or a cultural sight), its geographic location, and whether it comes recommended or not, as well as a unique identifier (in the form of a scrambled email address) for the author of the review.

Steve can use a task focused application called *Service X* to aggregate all the reviews on these sites that refer to places in New York and Boston. However, this still provides too much information; he only wants to read reviews written by members of his social network; people he knows and whose opinions he's interested in. Fortunately, using social network management application called *Service Y* Steve has defined the friends, family and colleagues that make up his social network and made this available online in a standard format that uniquely identifies people using the same scrambled email address method as the travel review sites. *Service X* can read this format and use Steve's definition of his social network to filter the travel reviews, only returning those written by people he knows.

With all the relevant reviews aggregated, Steve begins to plan his trip. He needs somewhere to stay when his flight arrives in New York, recommendations for places to eat, and suggestions for sights to see whilst in the city. The rest of the trip is unplanned, so as well as specific recommendations like these, he's also interested in general suggestions about how to spend his time in the region.

Using *Service X*, Steve attempts to locate information about accommodation in New York. He'll be travelling on a tight budget, so he looks specifically for backpacker hostels. The system returns a number of options that members of his network review negatively and recommend avoiding, and only one that is recommended. However, because *Service X* knows the room rates at these backpacker hostels (from the machine readable price lists on their web sites) and sees that they are comparable to those at budget guest houses, it also returns guest house reviews from Steve's social network on the basis that these might be suitable accommodation. In fact a guest house recommended by his colleague Kathy looks nice and he decides to reserve a room for his first few days.

Steve decides to leave New York plans for the time being and explore options for his trip to Boston and the surrounding countryside. He has heard very positive stories about the area but has no firm plans about what he'd like to do there or where exactly he'd like to go. Using *Service X* again, Steve explores what his social network has said about Boston. Unfortunately no one in his network has been there, or if they have then they haven't reviewed it. Having returned no results, *Service X* widens the scope by examining Steve's extended social network (not just members of his social network, but their friends, families, colleagues) and by looking at additional sources of information that may indicate knowledge of the area he wants to visit. Sure enough, his friend Holger has a sister Eva whose online CV says she studied at university in Boston; *Service X* reasons that she probably knows the surrounding area well. Steve has met Eva briefly in the past so contacts Holger and asks him to put them in touch so he can ask her advice about the best places to see and stay in that part of the world.

Specific research questions

The unifying research question given previously can be broken down into the following three questions, each of which will be addressed in one of the following sections below.

1. How do people use their social network for locating and exploring?

2. Does the semantic web provide a suitable platform for implementing a system capable of modelling the processes in question 1, capturing the social contextual information, and utilising it to facilitate these tasks?

3. What are the criteria that make such a system useful, usable, and appealing to users, and can these be fulfilled?

Summary of approach

The research will address these questions through a combination of exploratory study of first principles, technical implementation of a social context application, and evaluation of this application. Question 1 above will be addressed by a small scale user study to inform the model of social network usage that will be created in response to Question 2. Further technical development will investigate the issues of capturing the social context and relevant domain knowledge, and using these together specifically to facilitate the tasks of locating and exploring. Finally, in response to Question 3, an evaluation study will be carried out to gauge the success of the system in addressing user needs, relative to existing approaches. The research will be applied within the domain of travel, for reasons discussed below.

Research components and methodology

The following section further details the components of the research and how they will be carried out.

Research component 1
Pilot study of first principles: how social networks are used in locating and exploring

This small scale pilot study is designed to illuminate how people actually use their social networks, and provide an insight into the factors that influence this use. The study will take the form of both exploratory questions and testable hypotheses, with the findings informing the modelling of social context to be carried out in research component 2.

Exploratory questions

1. Why do people use their social networks to assist with locating and exploring? How significant are: the potential to receive a *personalised response*, the potential for direct access to *specific expertise*, and the *convenience* of asking someone nearby, in encouraging people to use their social network compared to other methods? Do other factors contribute?
2. Does knowing the opinions of the members of their networks help the individual to make better or different decisions?

These questions will likely be administered in questionnaire format, and responses will be analysed qualitatively as they will primarily be used to provide a broad overview of the issues.

Hypotheses

The following hypotheses will be used to investigate which specific factors influence how people use their social networks in locating and exploring. Specifically, the results of the tests of these hypotheses will allow us to understand and distinguish which factors people rely on when deciding whose opinion to seek about a particular problem, or whose knowledge to prioritise if several options are available. To test the hypotheses, participants will be asked to rate opinions of people who vary along the dimensions highlighted below. This will be administered using either a card sort exercise or a questionnaire and the results will be analysed quantitatively using tests of statistical significance.

Hypothesis A: Network proximity
H_1: the priority given to the opinion will increase with the network proximity of the nodes in the network
H_0: the priority given to the opinion will not be affected by the network proximity of the nodes in the network

In this instance network proximity is defined as the number of network "hops" that separate two people (nodes) in the social network.

Hypothesis B: Nature of relationship
H_1: the priority given to the opinion will vary depending on the nature of the social relationship between the nodes in the network
H_0: the priority given to the opinion will not vary depending on the nature of the social relationship between the nodes in the network

Whilst there is undoubtedly a correlation between certain types of social relationship (such as parent, sibling, spouse etc) and high network proximity, these two factors will be examined independently in the study.

Hypothesis C: Perceived domain expertise
H_1: the priority given to the opinion will increase with the perceived domain expertise of the node providing the opinion
H_0: the priority given to the opinion will not be affected by the perceived domain expertise of the node providing the opinion

It would be interesting and informative to investigate a number of other factors such as: the effect of the perceived value of the knowledge sought, on the distance people will travel in the network to obtain it. For example, will people use a more extended network when looking for employment opportunities than when looking for travel recommendations? However, due to the depth of these questions they fall outside the scope of this pilot study and could form the basis for further research.

Research component 2
Research into modelling, capturing, and utilising social context: a social context application

This component of the research will investigate the instantiation of the conceptual architecture shown above (Figure 1) in a social domain, with the tasks of locating and exploring. The application will attempt to address in an online setting the specific user need of accessing knowledge held by the user's social network.

System walkthrough

The processes to be embodied in the social context application consist of two subcomponents, population of the system and use of the system.

The population process is described below under system inputs and outputs. The use process is outlined in Figure 2 below and can be summarised in natural language as: "Has the social network defined by the user shared knowledge that answers the user's query? If not, who else does the user know who could be asked, and who might actually have an answer but just hasn't explicitly shared this knowledge yet."

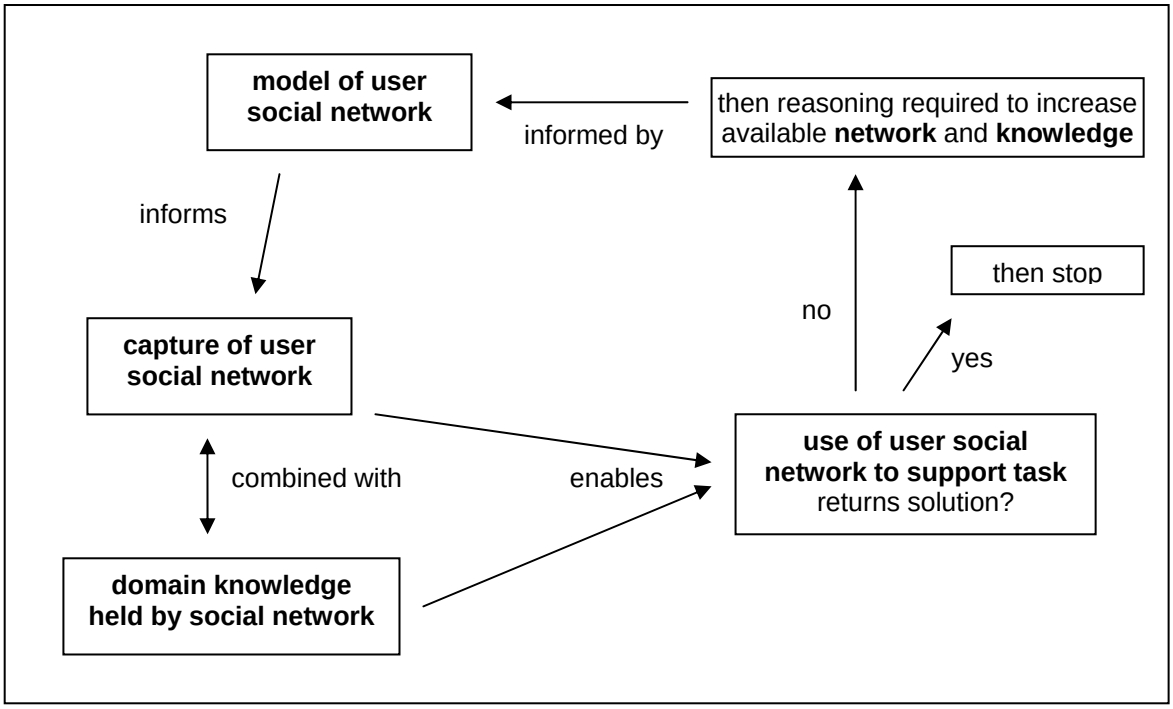

Figure 2: Cycle of using and reasoning about social network and the knowledge they hold, to support locating and exploring

In more formal terms, this process is made up of the following steps:

1. Does the social network the user has defined have knowledge about or a solution to the user's query?
2. If yes, then the system should provide the answer. This is a direct example of using the social network to facilitate task performance.
3. If not, then the system must reason along two dimensions:
    a. Who else the user may know that might have an answer, who hasn't been explicitly mentioned in the social network definition, but may be added for use by the application. For example, this might involve reasoning that a person is known to the user because they work in the same department, even though they weren't explicitly mentioned in the definition of the network.
    b. Who may have knowledge that is relevant to the query but that hasn't been expressed yet for use by the network. For example, this might involve reasoning that a person is likely to know about a city because they went to university there, even if they haven't explicitly expressed any opinions about the city.
4. The knowledge gained should then be used to provide an answer.
5. If several answers or solutions are available, then the system may use algorithms of how people use knowledge in a social network to reason about which advice to take in the context of the particular domain or task. This is a direct example of reasoning about the user's social network itself, with the goal of facilitating task performance.

Architectural components

The proposed application will consist of the following architectural components, depicted graphically in Figure 3:

- **Task-oriented application** for locating and exploring. Likely to be as simple as a web based query interface.

- **Context model**, defining how knowledge is shared in social networks, representing the relative importance to users of factors such as network proximity, nature of relationship, and perceived expertise.

- **Context knowledge base** holding a representation of the user's social network.

- **Domain model** such as a travel ontology in this case (see below for discussion of the use of the travel domain).

- **Domain knowledge base** of travel knowledge, populated by members of the network.

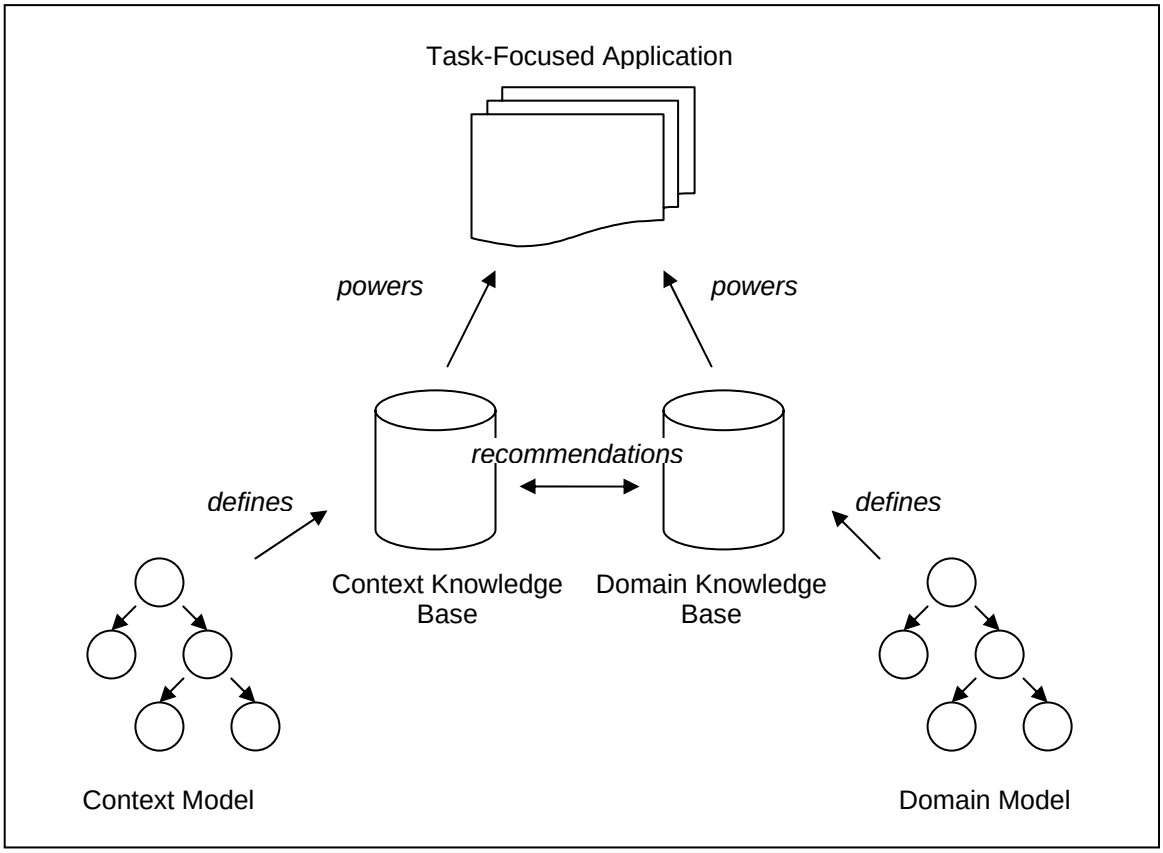

Figure 3: Architectural components of the social context application

Domain of the application: travel

The application will be deployed in the travel domain. Travel is seen as a suitable test domain for a number of reasons. Firstly, a key factor in the success of the proposed system is getting the engagement of potential users, and travel is a subject that people are motivated to talk about. It's common for people to discuss travelling in informal social situations, comparing experiences and sharing recommendations. It is also a domain where recommendations can be very valuable if a journey might be made a long way from home or in an unfamiliar culture, or if it's hard to gather information remotely before embarking on a trip.

Secondly, the domain is sufficiently complex to provide sufficient scope for the research. Many different types of information relevant to the locating and exploring tasks can be exchanged within the domain, from very concrete factual information supporting the locating task (such as recommended places to stay in New York) to more general background information about a destination to support the exploring task. The domain encompasses many different classes of objects (such as hotels, restaurants, cultural sights, journeys, transport

nodes, landscapes), each of which has different properties and may be complimented by a variety of resources (photos, fragments of information, textual accounts, maps, for example).

Thirdly, results obtained using the travel domain have the potential to generalise to other areas. The findings of this research should be applicable to any domain where word of mouth, opinion, and recommendation are important, where people are motivated to participate in sharing knowledge, and when the space of possible solutions is large and hard to manage. Other domains where the findings could be applicable or that could serve as useful test domains might include books, music, trades people such as plumbers or builders, or academic papers in a certain field. Furthermore, the classes of objects being considered have potential relevance at a local community level. Restaurant recommendations, for example, have as much relevance for people living in an area as for people travelling to it

Finally, many data sets of travel reviews already exist online that are largely non-semantic, but could be repurposed for this research and which could be complimented by the addition of social context. Examples include *OpenGuides*[10], *ChefMoz*[11], *Lonely Planet Thorn Tree*[12], and *Rough Guides*[13], a number of which are provided by companies publishing travel guide books. The ubiquity of the travel guide book medium validates that people are used to thinking of travel from this perspective. The proposed application would allow for a personalised, contextualised equivalent.

In addition, existing research has considered problems using the domain of travel as an example. This means there are previous findings to relate research outcomes to. For example, work by Toms, Freund, Kopak et al (2003) supports the difference between locating and exploring in a travel context but doesn't consider use of social context to support these tasks.

System inputs and outputs

To operate as designed the system would require the user to provide a **representation of their social network** as input (or have one generated for them from other sources) to populate the community knowledge base. It is assumed that whilst some users may provide their own hand-crafted definitions of their social network, the majority of users will rely on this definition being generated for them automatically. Possible solutions to this are discussed below.

Members of the social network will need to **annotate "travel objects"** (such as hotels, restaurants, places, sights etc) to populate the domain knowledge base. These annotations would include simply the URI of the object, a weighted recommendation of the object on a 5-point scale (highly recommended to highly not recommended), and any additional comments. These annotations would be provided manually (through a lightweight web-based mechanism), scraped from other sources, or both. They may also be complimented by generic resources from external sources in case the social network has no knowledge available on a given topic. The populated domain knowledge base may also contain machine-readable descriptions of the annotated travel objects, harvested directly in RDF (where available) from the websites of the travel objects or automatically generated from unstructured website text. These may be used to provide additional services to the user based on knowledge of the travel objects.

Performance of a task would be expected to start with an input from the user in the form of a **query** for information they would like to locate or a concept they would like to explore. The system would be expected to output possible **solutions to the user's query** in the form of knowledge gleaned from the network, plus proofs of **how the solution was reached**. Such proofs may include who a recommendation came from, how proximal they are in the network, in what way they are known, what their credentials and domain expertise are, etc. The exact

[10] http://openguides.org
[11] http://chefmoz.org
[12] http://thorntree.lonelyplanet.com
[13] http://roughguides.com

proofs returned would be informed by the results of the user study. These proofs would form the basis for rankings of the priority of results in the task focused application. In situations where no solution was found and the system has to reason about potential additional members of the network whose knowledge could be used, then these **new nodes** would be returned as output (for potential inclusion in the user's representation of their network), along with a **rationale** for how the new node was identified.

Methods of implementation and technologies used

Semantic web technologies

The use of semantic web technologies is central to this research component. Semantic web technologies will be used for their ability to describe resources in a distributed environment using agreed semantics, integrate these descriptions for a specific application, and then reason across them to provide added value. Conventional web technologies cannot provide the same degree of functionality and lend themselves more to systems that do not scale to diverse usage scenarios, and are overly centralised, requiring all users to subscribe to one system. Comparing the two approaches in the context of this application, semantic web technologies allow for users to manage their own definitions of their social network and their own annotations of travel objects, with these being aggregated by a third party system for querying, such as the application proposed here, or as illustrated in the motivating scenario provided earlier. In contrast, a conventional web approach would require users to all use one central system for the definition of their network, the publication of their annotations of travel objects, and for accessing this knowledge.

Several generic semantic web frameworks exist that could be utilised in development of this application. *Sesame*[14] is an "open source Java framework for storing, querying, and reasoning with RDF and RDF Schema", similarly *Jena*[15] (also in Java) is "a framework for building semantic web applications" that provides APIs for RDF and OWL, supports several different storage mechanisms, and includes a rule-based inference engine. Written in C, but with higher level interfaces to several other languages including PHP and Perl, the *Redland* application framework provides parsing, storage, and querying functionality for RDF. All of these frameworks could aid the management and use of the RDF descriptions of social networks and travel objects that are a fundamental to this application. Further analysis will be carried out to assess the most suitable framework or combinations to use.

Knowledge modelling and knowledge bases

A suitable knowledge modelling language is needed to represent the models of user social context and travel and knowledge base applications are required to support the populated knowledge bases. The most appropriate languages and technologies to use for these functions will be investigated.

Modelling and capturing social networks

The *Friend of a Friend* RDF vocabulary will be used as the foundation for defining user's social networks. Subject to the outcome of the user study, FOAF may need to be complimented by additional vocabularies if it is found that relations beyond the basic *knows* relation in FOAF are crucial in users making decisions about how to use their social network. A *Relationships* RDF vocabulary (discussed in Part I above) has been developed that attempts to describe a number of types of relationships between people, such as *friendOf*, *lostContactWith*, *knowsInPassing*, *knowsByReputation*, *worksWith*, *livesWith*, *neighborOf* etc. However, as Marshall and Shipman (2003) point out, it can be hard to "describe human oriented concepts" and this applies to the question of how to describe all the possible ways in which people know each other. For this reason the *Relationships* vocabulary is not seen a likely candidate for use in the application. If indeed people do attend to the nature of the relationship when deciding how to use knowledge from their network then a new approach will

[14] http://www.openrdf.org
[15] http://jena.sourceforge.net

need to be sought. In addition, it is not anticipated that the *Trust* vocabulary discussed by Golbeck and Hendler (2004) will be used, as it doesn't allow the context of the trust to be defined and it is predicted that this will be of significance to users when assessing the opinions of others.

Whilst FOAF data describing people and their social networks is already widely available on the semantic web, other users will be able to opt into the system by having a FOAF file generated for them by services such as *LiveJournal*[16], or by the use of applications that can mine sources such as email address books and produce a FOAF file for the user. Existing applications of this sort will be investigated for their suitability.

Several applications exist that can be used to collect and process existing FOAF data, as will be required by this application. To RDF as conventional web crawlers are to HTML, these so-called *scutters*[17] follow *rdfs:seeAlso* links in RDF files to gather semantic web data. Whilst these scutters can generally handle data from a wide range of RDF vocabularies, they have often originated in the FOAF community, ensuring they are well adapted to this application. To date, the *Slug*[18], *Elmo*[19] and *RDF Crawler*[20] scutters in Java, and the *AYF*[21] scutter in Perl, have been identified, with Slug having been installed and tested with promising results.

Representing travel knowledge and populating knowledge base

A comparison of existing travel and related ontologies and vocabularies is shown above in Table 3, within Part II of this report. Whilst a number of these could prove useful they are often too narrow in coverage, insufficiently granular in the necessary areas, or oriented towards the travel industry rather than the needs of the user in locating or exploring travel information. Where possible these existing ontologies will be reused, however a specific ontology representing user travel needs may need to be created focusing on simple travel objects such as hotels, restaurants, cultural sights, and destinations.

A number of existing data sources may be able to be repurposed to contribute to the population of the knowledge base, as described above in the discussion of the use of the travel domain. Such a process could make use of *GRDDL*[22], which provides a mechanism for transforming XML and XHTML into RDF, typically using algorithms represented in XSLT. Furthermore, the *Armadillo*[23] knowledge mining system, combined with a travel ontology, could be used for extracting semantic descriptions from the conventional (non-semantic) web sites of travel objects. The retrieval and analysis of such non-semantic web resources may also be aided by the use of web-mining tools such as *TextGarden*[24]. It is conceivable that generated semantic descriptions of conventional web resources could usefully be republished for use elsewhere, in which case tools such as *RAP* (RDF API for PHP)[25] could be used to parse, manipulate, serialize and serve up these RDF models. If generic resources need to be used to supplement the knowledge base at any time, this could take the form of calls to web services such as the *Google API*[26] to retrieve generic results on a particular topic.

Annotation and usage interfaces

An annotation interface would be required for users to populate the domain knowledge base with annotations of travel objects, as described above under *System inputs and outputs*. A

[16] http://www.livejournal.com
[17] http://rdfweb.org/topic/Scutter
[18] http://www.ldodds.com/blog/archives/000167.html
[19] http://www.openrdf.org/doc/elmo/users/index.html
[20] http://ontobroker.semanticweb.org/rdfcrawl/
[21] http://frot.org/rdfweb/ayf.html
[22] http://www.w3.org/2004/01/rdxh/spec
[23] http://www.dcs.shef.ac.uk/~alexiei/WebSite/University/Armadillo/
[24] http://kt.ijs.si/Dunja/textgarden/
[25] http://www.wiwiss.fu-berlin.de/suhl/bizer/rdfapi/
[26] http://www.google.com/apis/

customised version of the *Annotea*[27] server with a web-based client may be suitable for this purpose. The aim would be for a very simple interface in the style of the bookmarking system *del.icio.us*[28] which has acquired a significant number of users largely due to its simplicity of use and social features. Whilst a few RDF vocabularies exist that allow people to express liking for something (*ILike*[29]), or to express a review of something in a semantically meaningful way (*Review Vocabulary*[30]), these are often limited to certain types of objects and do not allow people to make recommendations explicitly. Consequently the recommendation mechanism described above under *System inputs and outputs* may need to be formalised as a simple ontology or vocabulary, or changes made to existing vocabularies to make them more suitable.

A query system would also be required to enable users to locate and explore travel information. It is anticipated that both these interfaces would be implemented as web-top applications using a server-side scripting language such as PHP.

Research component 3
Evaluation of the application

The following measures of success will underpin the evaluation of the social context application and together address both technical and human questions:

1. Is development of this application possible?
    a. Does the application work?
    b. Do the technology and methods used enable such a system to be implemented?
    c. If not, what are the limitations?
2. Does the application address and support the strategies people reported using?
3. Is it useful, usable, and desirable?
    a. Does the application give different results to those returned by a conventional web search for the same query?
    b. How do these results compare to what people would have suggested?

Contributions of the research

The research will make the following contributions

1. Results of the pilot study, illustrating which aspects of a social network are attended to when locating and exploring.

2. A representation of these findings of people use their social networks, in the form of an ontology

3. Algorithms for how to use the model for ranking the importance of results obtained from a network

4. A task focused and context aware application for locating and exploring

5. An architecture for contextualised task performance that's scalable across different contexts, tasks, and domains

6. A study evaluating the effectiveness of the application in meeting user needs

[27] http://www.w3.org/2001/Annotea/
[28] http://del.icio.us
[29] http://www.schemaweb.info/schema/SchemaDetails.aspx?id=239
[30] http://dannyayers.com/xmlns/rev/

Distinctiveness of the research

As demonstrated in the literature review, other approaches have been proposed that attempt to provide social context to the user in an online environment. Specifically other work has attempted to model and describe the nature of social relationships using semantic web technologies. These approaches are limited in being too narrow and insufficiently focused how they can support users in performing tasks. As discussed above, the trust vocabulary of Golbeck and Hendler (2004) doesn't specify the context in which a trust relationship exists between two people; for example a person may trust someone they know to lend them money if required, but not to recommend where to go on holiday. Therefore it is hard to use this vocabulary as the basis for reasoning that can support user tasks.

Mika (2004) has been significant in demonstrating how the semantic web can be mined for data about social relationships for use in social network analysis. However, the research proposed here is distinct in that it attempts to understand and model social networks so they can be used specifically to support user tasks online. The work of Chirita et al (2005) attempts to use the social context of documents or desktop objects to better support their manipulation and use in a semantic desktop. Consequently the social context belongs and is related to the object (who sent a particular email attachment to a recipient, for example) whereas the research proposed here is concerned with the social context of the user as the primary focus for supporting task performance.

Work plan

The research will be carried out between August 2005 and September 2007, giving a total of 26 months. The primary tasks consist of:

- User study, execution and analysis
    - August 2005 - **September 2005**
    - No dependencies

- Application development
    - August 2005 - **July 2006**
    - Dependent on outcome of user study, although some aspects can begin immediately
    - Dependent on development of knowledge modelling skills (beginning with attendance at SSSW, Spain, July 2005)

- Evaluation
    - October 2006 - **December 2006**
    - Dependent on creation of working system

- Writing dissertation
    - January 2007 - **June 2007**
    - Dependent on all previous stages

- Review and revision of dissertation
    - July 2007 - **September 2007**
    - Dependent on writing of dissertation

- Submission of dissertation
    - **September 2007**

- Writing papers and attending conferences
    - Ongoing. Target conference includes:
        - ISWC2005 Workshops (July and August 2005)

- IUI2006 (19 September 2005)
- WWW2006 (4 November 2005)
- ESWC2006 (28 November 2005)
- ISWC2006 (~May 2006)
- HCI and Pervasive computing conferences will also be investigated, as will conferences in 2007 as submission dates are released.

Milestone dates are shown in bold. 2 months contingency time is allocated between the end of July and the start of October 2006.

References

Almeida, R. B. and Almeida, V. A. F. (2004) A Community-Aware Search Engine. In *13th International Conference on World Wide Web (WWW2004)* ACM Press, New York, USA, pp. 413-421.

Berners-Lee, T. (1998) Semantic Web Road Map. World Wide Web Consortium (W3C).

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American,* 284, 34-43.

Brickley, D. and Miller, L. (2004) FOAF Vocabulary Specification.

Broder, A. (2002) A Taxonomy of Web Search. *ACM SIGIR Forum,* 36, 3-10.

Bush, V. (1945) As We May Think. *The Atlantic Monthly*.

Byrne, M. D., John, B. E., Wehrle, N. S. and Crow, D. C. (1999) The Tangled Web We Wove: A Taskonomy of WWW Use. In *SIGCHI Conference on Human Factors in Computing Systems (CHI1999)*.

Celma, O., Ramirez, M. and Herrera, P. (2005) Getting Music Recommendations and Filtering Newsfeeds From FOAF Descriptions. In *ESWC2005 Workshop on Scripting for the Semantic Web* Heraklion, Crete.

Chakrabarti, S., Srivastava, S., Subramanyam, M. and Tiwari, M. (2000) Using Memex to Archive and Mine Community Web Browsing Experience. *Computer Networks,* 33, 669-684.

Chirita, P. A., Gavriloaie, R., Ghita, S., Nejdl, W. and Paiu, R. (2005) Activity Based Metadata for Semantic Desktop Search. In *2nd European Semantic Web Conference (ESWC2005)* Heraklion, Crete.

Choo, C. W., Detlor, B. and Turnbull, D. (1999) Information Seeking on the Web: An Integrated Model of Browsing and Searching. In *62nd Annual Meeting of the American Society for Information Science* Washington DC, USA, pp. 3-16.

Cockburn, A. and McKenzie, B. (2000) What Do Web Users Do? An Empirical Analysis of Web Use. *International Journal of Human-Computer Studies*.

Dey, A. K. (2001) Understanding and Using Context. *Personal and Ubiquitous Computing,* 5, 4-7.

Dey, A. K., Abowd, G. D. and Salber, D. (2001) A Conceptual Framework and a Toolkit for Supporting Prototyping of Context-Aware Applications. *Human-Computer Interaction,* 16, 97-166.

Dieberger, A., Dourish, P., Hook, K., Resnick, P. and Wexeblat, A. (2000) Social Navigation: Techniques for Building More Usable Systems. *Interactions*, 37-45.

Dieberger, A., Hook, K., Svensson, M. and Lonnqvist, P. (2001) Social Navigation Research Agenda. In *SIGCHI Conference on Human Factors in Computing Systems (CHI2001)* ACM Press, Seattle, USA, pp. 107-108.

Ding, L., Finin, T. and Joshi, A. (2005) Analyzing Social Networks on the Semantic Web. *IEEE Intelligent Systems*.

Domingue, J. B., Dzbor, M. and Motta, E. (2003) Semantic Layering with Magpie. In *Handbook on Ontologies in Information Systems* Springer Verlag.

Domingue, J. B., Dzbor, M. and Motta, E. (2004) Magpie: Supporting Browsing and Navigating on the Semantic Web. In *Conference on Intelligent User Interfaces (IUI2004)* Portugal.

Dourish, P. (2001) Seeking a Foundation for Context-Aware Computing. *Human-Computer Interaction,* 16, 229-241.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D. C. (2003) Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *26th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'03)* ACM Press, Toronto, Canada.

Duri, S., Cole, A., Munson, J. and Christensen, J. (2001) An Approach to Providing a Seamless End-User Experience for Location-Aware Applications. In *WMC '01* Rome, Italy.

Emmanouilides, C. and Hammond, K. (2000) Internet Usage: Predictors of Active Users and Frequency of Use. *Journal of Interactive Marketing,* 14, 17-32.

Erickson, T. and Kellogg, W. A. (2000) Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction,* 7, 59-83.

Fischer, G. (2001) Articulating the Task at Hand and Making Information Relevant to It. *Human-Computer Interaction,* 16, 243-256.

Gil, Y. and Ratnakar, V. (2002) Trusting Information Sources One Citizen at a Time. In *1st International Semantic Web Conference (ISWC2002)*.

Golbeck, J. and Hendler, J. (2004) Reputation Network Analysis for Email Filtering. In *1st Conference on Email and Anti-Spam (CEAS2004)*.

Granovetter, M. S. (1973) The Strength of Weak Ties. *The American Journal of Sociology,* 78, 1360-1380.

Gravano, L., Hatzivassiloglou, V. and Lichtenstein, R. (2003) Categorizing Web Queries According to Geographical Locality. In *CIKM '03* New Orleans, USA.

Gray, E., Seigneur, J.-M., Chen, Y. and Jensen, C. (2003) Trust Propagation in Small Worlds. In *Trust Management 2003*.

Grudin, J. (1994) Groupware and Social Dynamics: Eight Challenges for Developers. *Communications of the ACM,* 37, 92-105.

Grudin, J. (2001) Desituating Action: Digital Representation of Context. *Human-Computer Interaction,* 16, 269-286.

Guha, R., McCool, R. and Fikes, R. (2004) Contexts for the Semantic Web. In *3rd International Semantic Web Conference (ISWC2004)*.

Guha, R., McCool, R. and Miller, E. (2003) Semantic Search. In *Twelfth International Conference on World Wide Web (WWW2003)* ACM Press, Budapest, Hungary, pp. 700-709.

Hansen, F. A., Bouvin, N. O., Christensen, B. G., Gronbaek, K., Pedersen, T. B. and Gagach, J. (2004) Integrating the Web and the World: Contextual Trails on the Move. In *HT '04* Santa Cruz, USA.

Heath, T., Dzbor, M. and Motta, E. (2005) Supporting User Tasks and Context: Challenges for Semantic Web Research. In *ESWC2005 Workshop on End-User Aspects of the Semantic Web (UserSWeb)* Heraklion, Crete.

Heath, T., Motta, E. and Dzbor, M. (2005) Uses of Contextual Information to Support Online Tasks. In *1st Doctoral Symposium of the Advanced Knowledge Technologies Project* Milton Keynes, UK.

Kalfoglou, Y., Alani, H., Schorlemmer, M. and Walton, C. (2004) On the Emergent Semantic Web and Overlooked Issues. In *3rd International Semantic Web Conference (ISWC2004)*.

Khedr, M. and Karmouch, A. (2004) Negotiating Context Information in Context-Aware Systems. *IEEE Data Engineering Bulletin*, 21-29.

Lawrence, S. (2000) Context in Web Search. *IEEE Data Engineering Bulletin,* 23, 25-32.

Leake, D. B. and Scherle, R. (2001) Towards Context-based Search Engine Selection. In *6th International Conference on Intelligent User Interfaces* ACM Press, Santa Fe, USA, pp. 109-112.

Linden, G., Smith, B. and York, J. (2003) Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Data Engineering Bulletin*, 76-80.

Marshall, C. C. and Shipman, F. M. (2003) Which Semantic Web? In *Fourteenth ACM Conference on Hypertext and Hypermedia (HT'03)* ACM Press, Nottingham, United Kingdom, pp. 57 - 66.

McBride, B. (2002) Four Steps Towards the Widespread Adoption of a Semantic Web. In *1st International Semantic Web Conference (ISWC2002)*.

Middleton, S., De Roure, D. and Shadbolt, N. (2001) Capturing knowledge of user preferences: ontologies in recommender systems. In *International Conference on Knowledge Capture (K-CAP2001)* Victoria, Canada.

Mika, P. (2004) Bootstrapping the FOAF-Web: An Experiment in Social Network Mining. In *First International Workshop on FOAF, Social Networks and the Semantic Web* Galway, Ireland.

Morrison, J. B., Pirolli, P. and Card, S. K. (2001) A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. In *SIGCHI Conference on Human Factors in Computing Systems (CHI2001)* ACM Press, Seattle, USA, pp. 163-164.

Motta, E., Buckingham Shum, S. and Domingue, J. (2000) Ontology-Driven Document Enrichment: Principles, Tools and Applications. *International Journal of Human-Computer Studies,* 52, 1071-1109.

Olston, C. and Chi, E. H. (2003) ScentTrails: Integrating Browsing and Searching on the Web. *ACM Transactions on Computer-Human Interaction (TOCHI),* 10, 177-197.

Oulasvirta, A. (2004) Finding Meaningful Uses for Context-Aware Technologies: The Humanistic Research Strategy. In *SIGCHI Conference on Human factors in Computing Systems (CHI2004)* ACM Press, Vienna, Austria.

Pirolli, P. and Card, S. K. (1998) Information Foraging Models of Browsers for Very Large Document Spaces. In *Working Conference on Advanced Visual Interfaces (AVI'98)* ACM Press, L'Aquila, Italy, pp. 83-93.

Quan, D. and Karger, D. R. (2004) How to Make a Semantic Web Browser. In *13th International Conference on World Wide Web (WWW2004)* ACM Press, New York, USA, pp. 255-265.

Reber, A. S. (1997) *The Penguin Dictionary of Psychology,* Penguin Books Ltd.

Richardson, M., Agrawal, R. and Domingos, P. (2003) Trust Management for the Semantic Web. In *2nd International Semantic Web Conference (ISWC2003)* Florida, USA.

Rose, D. E. and Levinson, D. (2004) Understanding User Goals in Web Search. In *13th International Conference on World Wide Web (WWW2004)* ACM Press, New York, USA, pp. 13-19.

Schraefel, M. C., Shadbolt, N. R., Gibbins, N., Harris, S. and Glaser, H. (2004) CS AKTive Space: Representing Computer Science in the Semantic Web. In *13th International Conference on World Wide Web (WWW2004)* ACM Press, New York, USA, pp. 384-392.

Schuler, D. (1994) Social Computing. *Communications of the ACM,* 37, 29.

Sellen, A. J., Murphy, R. and Shaw, K. L. (2002) How Knowledge Workers Use the Web. In *SIGCHI Conference on Human Factors in Computing Systems (CHI2002)* ACM Press, Minneapolis, USA, pp. 227-234.

Shardanand, U. and Maes, P. (1995) Social Informatin Filtering: Algorithms for Automating "Word of Mouth". In *CHI'95*.

Svanaes, D. (2001) Context-Aware Technology: A Phenomenological Perspective. *Human-Computer Interaction,* 16, 379-400.

Svensson, M., Hook, K., Laaksolahti, J. and Waern, A. (2001) Social Navigation of Food Recipes. In *SIGCHI'01* Seattle, USA.

Tauscher, L. and Greenberg, S. (1997) How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human-Computer Studies,* 47, 97-137.

Toms, E. G., Freund, L., Kopak, R. and Bartlett, J. C. (2003) The Effect of Task Domain on Search. In *2003 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON2003)* IBM Press, Toronto, Canada, pp. 303-312.

Voss, A. and Kreifelts, T. (1997) SOaP: Social Agents Providing People with Useful Information. In *GROUP'97* Phoenix, USA.