Uses of contextual information to support online tasks

Conference or Workshop Item

Uses of Contextual Information to Support Online Tasks

Tom Heath, Enrico Motta, Martin Dzbor

Knowledge Media Institute, The Open University,
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
{t.heath, e.motta, m.dzbor}@open.ac.uk

Abstract. In this paper we make a case for the role of user context information in supporting task performance online, examine previous attempts at representing and making use of user context factors, and highlight the limitations of existing tools and services. We then suggest how the emergent Semantic Web might be able to better facilitate the capture of knowledge regarding user context, and provide the means for its reuse in supporting the performance of tasks online.

1 Introduction

Many types of activities and tasks have been identified that users carry out online, and these have been classified in a number of different ways. For example, Broder [1] identifies three different types of web search: *navigational*, *informational*, and *transactional*. Also assuming an information need as the basis for an information search, Morrison, Pirolli and Card [2] focus on the web activities that significantly affect the actions people take and the decisions they make. They classify these according to the variables *purpose*, *method*, and *content*. Placing a stronger emphasis on the broader goals of the user, and considering tasks that are not purely driven by an information need, Sellen, Murphy and Shaw [3] identify six tasks that users perform on the web. Addressing all internet platforms rather than just the web, Heath, Dzbor and Motta [4] maintain this task-focused approach and identify ten tasks users perform online.

In addition to addressing the full range and diversity of the tasks users perform online, it is essential to recognise that these tasks aren't performed in isolation. For example, one of the ten online tasks identified by Heath et al [4] is *arranging*. They illustrate this task with the scenario of an internet user making arrangements online to attend an international conference. The task in the scenario is complex and challenging using existing internet tools such as travel web sites, and the user relies on the knowledge and past experiences of a colleague to complete the task whilst satisfying his preference for low cost over convenience.

Additional factors not mentioned in the scenario could also be seen to impact upon the performance of the task, such as preferences the user has about accommodation or the time they currently have available for making the arrangements. These contextual factors can be seen to shape or define the nature of the task and to aid in its

performance. However, despite this, even the most widely used internet tools such as search engines and web browsers are not well adapted to making use of them.

2 Support for Context in Existing Internet Tools

2.1 Context in Web Search

Traditionally conventional search engines have treated all users exactly the same [5] [6] 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[1], Ask Jeeves[2], 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 [7] should enable these kinds of statements to be made in a way that is meaningful to both humans and machines.

Previous research has attempted to improve support for context in domains such as web searching, but addressing the topical context of the search term, not the context of the user. For example, Lawrence [5] 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 [8] 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-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 domain- and task-specific services. This would suggest that making use of knowledge related to the user's context, rather than

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

specific to a particular task or domain, would actually be of most value in assisting online task performance.

2.2 Exploiting Knowledge of User Context

Whilst focused purely on one tool, the discussion above suggests that if users are to be better supported in performing tasks online, such as those described by Heath et al [4], internet tools need to be able to make better use of knowledge about aspects of user context. In the same paper, the authors [4] briefly introduce six aspects of a user's context which may be pertinent when performing tasks online, as they would be in the offline world: a user's *social networks*, their *previous experiences*, *preferences* they hold, their *current location*, services or third parties they *trust*, or the *resources* they have available for performing the task.

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; 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.

Existing attempts to support these user context factors are discussed below. Whilst these primarily take the form of web tools or services, we follow the rationale of Heath et al [4] that all internet platforms should be considered, not just the web.

One service that has made effective use of community and social factors to help people perform tasks is the online retailer Amazon[4]. As well as keeping a record of your purchase history and using this to recommend additional items you might like to buy, the service uses a collaborative filtering mechanism to suggest items that others who match your shopping profile have bought. Not only can this be used in locating specific items, it also supports the *grazing* task identified in [heath et al]; 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, so the context is restricted to that one system.

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 context factor 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 [9]. 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

[44] http://www.amazon.com

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 et al [10] 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 et al [11] 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 [12]. 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" [13]. 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 et al [14] 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 [15] 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".

A number of attempts have been made to create trust or reputation systems on the web. 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 office, 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, despite the importance of trusted third parties in contributing to a user's context. Some well-known organisations have attempted to set up schemes for validating online retailers to demonstrate their trustworthiness. However, the criteria used to assess the trustworthiness are often not clear to the user and likely carry less weight than validations obtained from highly trusted sources such as friends or family. Online auction services such as eBay 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 recommendations, in that the knowledge held by one site cannot be used by others.

Within the field of Computer Supported Collaborative Working (CSCW), Voss and Kreifelts [16] describe a system that takes attempts to take *community*, *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 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 [17] suggest that people will make use of strategies that minimise the resources required in locating information, supporting the existence of systems such as these.

3 Limitations of Existing Approaches

Whilst the approaches outlined above undoubtedly have a contribution to make in representing user context and using it to assist in task performance online, certain limitations regarding capture and reuse of contextual information are common to them all.

At present, if contextual information is to be taken into consideration when performing tasks online, the user must be aware of its significance and be able to explicitly convey it to each tool or service used in performance of a task. For example, current online trust or reputation systems may suggest who a user should trust but do not necessarily allow a user to convey which third parties they already do trust to assist in performance of the task.

In many cases, the capture and use of knowledge about user context falls solely to the user themselves, and this process must be repeated whenever the knowledge needs to be reused. In other cases, use of a particular tool or service may create contextual information which would be desirable to reuse in performing another task using a different tool or service.

This use and subsequent reuse of contextual information is not always possible using the tools currently available. For example, whilst "most problem-solving tasks are collaborative in nature" [18] (page 271), current search engines do not provide the means for users to locate information with assistance from others. Returning to the

example of Amazon referred to earlier, substantial data is held about a customer's purchase history, a rich source of contextual knowledge. However, the user has no way of automatically exposing this for consumption by other services. For example, it is not currently possible for a user to receive automatic film recommendations from other parties they choose, based on their history of DVD purchases from Amazon. Services such as these operate as so-called *closed-worlds*, where only parties approved by the service have access to the user's contextual information. Where data sharing arrangements do exist, the formats and protocols for exchanging the data are often negotiated on a case by case basis rather than according to common standards.

This reliance on closed worlds, where contextual information cannot be transferred across situations to assist in performance of other tasks, can be seen to create an artificial distinction between task performance offline and online. In the offline world, contextual knowledge is used regularly and intuitively, whilst this is frequently not possible online. This distinction highlights the limited *social translucence* [19] of current internet tools.

4 Conclusions and Future Research

If internet tools and services are to make better use of user context factors in assisting with the performance of tasks online certain developments need to take place, and Semantic Web technologies may provide the technical infrastructure required to do so.

Users must have better means to capture, create, manage, share, and reuse knowledge about their context, and this must apply to all the factors identified in [4]. This is in turn dependent on the existence of agreed formats for describing each of these context factors. Significant progress is already being made in this area. For example, the *Friend of a Friend* (FOAF) RDF vocabulary [20] provides a means for people to describe the social networks they belong to. This may be supplemented with more granular vocabularies defining trust relationships [21]. A number of means also exist for people to express their location in semantically meaningful ways. Privacy is clearly crucial to this process, therefore it is essential that the user remains in control of their data and can determine which trusted parties to share it with.

To build on these developments, task-focused internet applications are required that can make use of the increasing amounts of contextual information available, and use this to better support users in performing tasks online.

Achieving this is dependent on certain architectural changes taking place online. Supporting the tasks outlined by Heath at al [4] whilst taking greater account of user context cannot be achieved simply by building larger and more powerful closed systems. No system, however powerful, can address the needs of all tasks and make use of the full range of contextual information. A more distributed architecture is required, where knowledge can be consumed from a wide range of locations and sources, and integrated meaningfully. The Semantic Web provides the means for this kind of integration to be carried out by machines as well as by humans, hopefully resulting in a more contextualised user experience online.

References

1. Broder, A.: A Taxonomy of Web Search. ACM SIGIR Forum 36 (2002) 3-10
2. Morrison, J. B., Pirolli, P., Card, S. K.: A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. In: Proc. Conference on Human Factors in Computing Systems (CHI'01) (2001) 163-164
3. Sellen, A. J., Murphy, R., Shaw, K. L.: How Knowledge Workers Use the Web. In: Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI2002) (2002) 227-234
4. Heath, T., Dzbor, M., Motta, E.: Supporting User Tasks and Context: Challenges for Semantic Web Research. In: Proc. ESWC UserSWeb Workshop on User Aspects of the Semantic Web (2005)
5. Lawrence, S.: Context in Web Search. IEEE Data Engineering Bulletin 23 (2000) 25-32
6. Almeida, R. B., Almeida, V. A. F.: A Community-Aware Search Engine. In: Proc. 13th International Conference on World Wide Web (WWW2004) (2004) 413-421
7. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284 (2001) 34-43
8. Leake, D. B., Scherle, R.: Towards Context-based Search Engine Selection. In: Proc. 6th International Conference on Intelligent User Interfaces (2001) 109-112
9. Tauscher, L., Greenberg, S.: How people revisit web pages: empirical findings and implications for the design of history systems. International Journal of Human-Computer Studies 47 (1997) 97-137
10. Chakrabarti, S., Srivastava, S., Subramanyam, M., Tiwari, M.: Using Memex to Archive and Mine Community Web Browsing Experience. Computer Networks 33 (2000) 669-684
11. Motta, E., Buckingham Shum, S., Domingue, J.: Ontology-Driven Document Enrichment: Principles, Tools and Applications. International Journal of Human-Computer Studies 52 (2000) 1071-1109
12. Domingue, J. B., Dzbor, M., Motta, E.: Semantic Layering with Magpie. In: S. Staab and R. Studer, (eds.): Handbook on Ontologies in Information Systems. Springer Verlag (2003)
13. Domingue, J. B., Dzbor, M., Motta, E.: Magpie: Supporting Browsing and Navigating on the Semantic Web. In: Proc. Conference on Intelligent User Interfaces (IUI2004) (2004)
14. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D. C.: Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In: Proc. 26th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'03) (2003)
15. Bush, V.: As We May Think. The Atlantic Monthly (1945)
16. Voss, A., Kreifelts, T.: SOaP: Social Agents Providing People with Useful Information. In: Proc. GROUP'97 (1997)
17. Pirolli, P., Card, S. K.: Information Foraging. Psychological Review 106 (1999) 643-675
18. Wexelblat, A., Maes, P.: Footprints: History-Rich Tools for Information Foraging. In: Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'99) (1999) 270-277
19. Erickson, T., Kellogg, W. A.: Social Translucence: An Approach to Designing Systems that Support Social Processes. ACM Transactions on Computer-Human Interaction 7 (2000) 59-83
20. Brickley, D., Miller, L.: FOAF Vocabulary Specification.http://xmlns.com/foaf/0.1/ (2004)
21. Golbeck, J., Parsia, B., Hendler, J.: Trust Networks on the Semantic Web. In: Proc. Cooperative Intelligent Agents (2003)