Supporting serendipitous discovery

R.S.Aylett[(1)], D.S.Bental[(1)], R. Stewart[(1)], J. Forth[(2)] and G.Wiggins[(2)]

[1]Heriot-Watt University
Edinburgh EH14 4AS, UK
{r.s.aylett, d.s.bental, r.stewart} @hw.ac.uk

[2]Queen Mary University of London
London E1 4NS, UK
{jamie.forth, geraint.wiggins} @qmul.ac.uk

ABSTRACT

This paper discusses the problem of supporting serendipitous discovery in the age of the internet. It describes a mobile-based Personal Assistant, SerenA, designed to help produce unexpected and valuable connections between a researcher and other researchers as well as between a researcher and ideas, resources and events.

1. INTRODUCTION

Many researchers report memorable experiences in which something fortunate – a person, an artifact, a person – is encountered accidentally, in the sense that it was not being explicitly sought. The discovery of penicillin is an often-cited instance, but less spectacular examples may include browsing in a library and coming upon a book unexpectedly, or meeting a person in a non-work related setting such as a party or an airport. The literature suggests that not only are serendipitous experiences relatively common in research, but that they contribute to the generation of new knowledge [2].

However in the age of the internet, query-based web search has become a leading source of information for researchers. This supposes that users have well-defined notions of the resources they seek, and the system's role is to facilitate efficient retrieval. We argue that this has reduced the opportunities for serendipitous discovery [5]. Recommender systems using collaborative filtering may suggest what others in the field have found useful, typically relying on matched user profiles to determine likely relevant resources. However, retrieving resources because they are valued by other "similar" users is unlikely to result in serendipitous material.

While, almost by definition, serendipitous connections cannot be 'automatically' generated, the SerenA project[1] is building a Research Assistant system to provide recommendations for resources that adapt to the different interests, goals and needs of its users. These resources may include people to talk to, papers to read [6], or events to attend. A first prototype will be available on mobile devices, with a focus on creating unexpected links between people that are relevant, timely and sensitive to location.

2. ARCHITECTURE

In order to make interesting and unexpected connections, SerenA explores and combines information from many different sources, using the growth of semantic web resources. In particular, increasing amounts of information from different domains are being made available as Linked Open Data(LOD). LOD uses emerging syntactic and semantic standards such as The Resource Description Framework (RDF) and OWL, and is available for query across the web. Serena also relies on the Semantic Web ontologies that are currently being developed and integrated to express information in different domains, such as FOAF[1] to describe people and relationships, DBpedia[2] for general knowledge, GeoNames[3] for geographic locations and Dublin Core[4] for publications. These ontologies are being realised in individual web-accessible databases which can be searched by tools such as Sindice[5] or merged into larger databases of machine-readable information such as FactForge[6]. As a result of these initiatives it is now possible to combine information from many different sources, for example to link location data obtained from a mobile device to information about nearby places of interest.

Figure 1 shows an agent-based architecture within which such diverse sources of semantically-marked information can be integrated. SerenA agents fall into two broad classes. Server agents perform specific tasks at the request of other agents, such as downloading information from the Web, analysing free-text documents, or looking up events in a user's calendar. Autonomous agents are capable of sophisticated proactive behaviour based on their specialised goals, reasoning capabilities, and the information currently available. JADE has been used as an agent framework.

Information is passed between agents to complete short- to medium-term goals. Agents can also hold persistent beliefs, to satisfy longer-term goals. Both domain knowledge and agent control commands are represented in RDF, a good choice for expressing highly structured knowledge- based information. The message envelopes, wrapping agent RDF-based communications, meet the FIPA Agent Communication Language specification[7]. This approach eliminates the need to translate domain knowledge acquired from the Semantic Web to internal agent knowledge representations. It can also support multiple levels of agent reasoning.

The SerenA ontology is currently divided into four sub-ontologies: *system* (agent commands and system events); *user-data*; *goals* (inferred from user-data); and *suggestions* (types of connections and suggestions to be generated).

---

[1] http://www.serena.ac.uk

[1] http://www.foaf-project.org

[2] wiki.dbpedia.org

[3] www.geonames.org

[4] dublincore.org

[5] sindice.com

[6] factforge.net

[7] http://www.fipa.org/repository/aclspecs.html

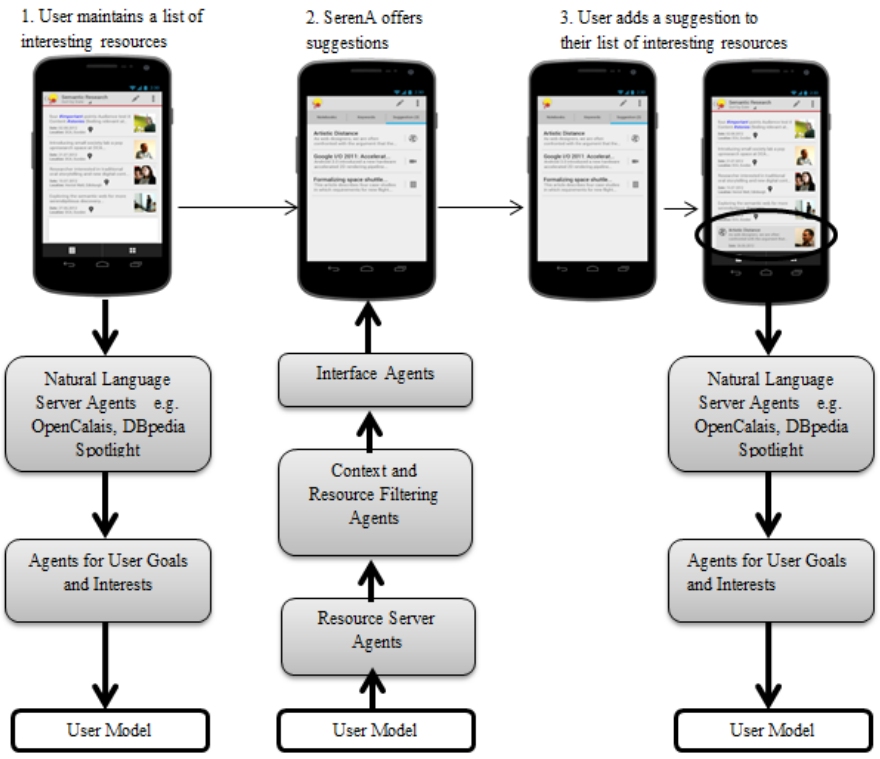

Figure 1 Serena Archiecture

User devices – Android-based mobiles in the current prototype – access the back-end agent system via a Websocket endpoint, a simple and stable server independent of JADE. This gives clean decoupling between user front-end and agent back-end and allows the Android app through which users interact to remain functional if the agent system goes down.

3. SERENA USERS

While public information is gathered from the Semantic Web, individual information is also gathered (with agreement) from SerenA's users. The user model itself uses RDF and standard ontologies such as the Cognitive Characteristics Ontology[8]. Rather than asking users to list their interests explicitly, SerenA infers users' likely interests from a range of sources. Users can tag web sites that they find of interest; key concepts are derived from an automated natural language analysis of those websites, and these concepts are added to the user model. Where users have published papers, inferences may be made about their research interests from the metadata attached to those papers and available in databases such as DBLP[9]. Users may also give more detail by supplying texts, such as a thesis proposal or personal web pages for automated analysis. The user's feedback on SerenA's own suggestions is added to the user model. Networks of existing contacts with other researchers can be inferred from publications databases, and included in the user model. SerenA balances search and recommendation techniques so that the model of the user's interests, goals and needs is gathered over a period and is used to extend queries and to provide recommendations. Up-to-date information is gathered from automated analysis of the user's tweets, and immediate location information from the user's mobile device.

It is recognized that explanation mechanisms are needed in systems that recommend resources to users [3, 10]. In order for users to trust SerenA's recommendations and continue to use SerenA, various kinds of explanation are needed [11]. Information is taken from multiple sources and combined, so that it will be important to track provenance. Different sources of information have different degrees of reliability. The user model itself must be scrutable as inferences about users' interests may be incorrect or out of date, and users will need to understand the basis on which recommendations are made and be able to correct mistaken inferences in the user model.

Initial studies were undertaken within SerenA to identify requirements for technologies which could facilitate serendipity. These studies identified a range of system functions, usability and social and contextual conditions [9]. Users wanted a system that could engage them with the information on offer, stimulate their curiosity and maintain their interest. In particular, users felt that it was important that a system could highlight the significance of a discovery and its relevance to their own interests and goals.

Serendipitous discoveries typically involve making connections which are both unexpected and valuable, and they require insight on the part of the user [5]. Unexpected information is often especially interesting and requires further explanation [1] and this is all the more important for potentially serendipitous connections, which may be missed because unlike other recommendations, the value of a serendipitous connection may not be immediately obvious. Serendipitous connections are made across different domains of interest. They may connect apparently unrelated ideas or researchers who work in different disciplines. A solution to a problem may be encountered serendipitously while

[8]http://smiy.sourceforge.net/cco/spec/cognitivecharacteristics.html

[9]http://dblp.uni-trier.de/

exploring a different problem. When people and ideas that could be useful in one context are encountered in a very different context, their significance may be missed altogether. It is important to highlight the potential significance of the connection to avoid this. Conversely, connections across very disparate domains are likely to carry a high risk of failure. Users must have enough information to judge whether a connection is likely to be valuable before they put a large amount of effort into following it up. In particular, if people are advised to connect with other researchers they are likely to need good reasons to do so before potentially wasting their own and other people's time.

Our initial prototype presents suggestions as a simple list (adhering to mobile platform design guidelines and best practice), without providing an underlying explanation to the user regarding why the suggestion was made, and leaving it up to them to explore and act upon the suggestions using web tools such as their browser. In more fully functioning prototypes the presentation of the suggestions and overall user experience will be explored through 'Classic' and 'Delightful' interface designs. These draw on psychological research which links emotional states, such as open-mindedness, to the individual's creative process. Russ [7] describes how affective states influence divergent thinking and transformational abilities. She specifically talks about openness, affective pleasure in challenge and problem solving, and cognitive integration. She has found that some emotions can broaden search and that mood-relevant cognition "triggers a broad associative network" highly relevant to making serendipitous connections.

Our evaluation methods will include implicit feedback collection and logging, with interviews using the feedback and logs to ascertain how much 'serendipity' users perceive in the connections they have made through SerenA. Our mobile studies will allow us to make use of a contextual testing environment which shifts away from the artificiality of a computer lab, and can make full use of the bundled data recording capabilities of mobile devices (for example, user-aware logging, image capture, note taking, and location detection). The inherent native functionality of mobile devices can of course be used not only for evaluation purposes, but may be harnessed in the conceptualization and development of the 'Delightful' interface, for example locative capabilities (e.g. GPS, wifi, digital compass), connection technologies (e.g. Bluetooth, NFC) and gestural input (e.g. multi-touch screens, accelerometers, gyroscopes and sensors).

4. CONCLUSIONS

SerenA has created an agent-based approach to the semantic web, allowing added value to be gained from the configuration and connection of the growing set of semantically-marked resources, with full expandibility as new such resources become available. It has combined this with a user-centred design approach [9] in which a serendipity diary application has allowed the collection and analysis of serendipitous occurrences.

An initial mobile prototype has demonstrated the feasibility of location-sensitive generation of connections between researchers and more formal evaluation is about to commence.

5. ACKNOWLEDGMENTS

We would like to acknowledge the work of SerenA colleagues at Dundee University (Mel Woods, Debbie Maxwell and Jamie Shek), UCL (Ann Blandford and Stephann Makri) and Nottingham University (Sarah Sharples, Xu Sun and Genovefa Kefalidou) for work on analysing serendipity and developing user requirements and evaluation approaches. This research is supported by the RCUK-funded SerenA Project EP/H042741/1.

6. REFERENCES

[1] Byrne, E. (2006) A Logical Framework for Identifying and Explaining Unexpected News. Computing and Informatics Vol 25 pp 127—152.

[2] Foster, A. and Ford, N. (2003). Serendipity and Information Seeking: An Empirical Study. Journal of Documentation, 59(3), pp. 321-340

[3] Herlocker, J., Konstan, J. and Riedl, J. 2000. Explaining Collaborative Filtering Recommendations. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, Pennsylvania, USA, 241-250.

[4] D. Liestman. Chance in the midst of design: approaches to library research serendipity. RQ, 31(Summer 1992):524–532, 1992.

[5] Makri, S. & Blandford, A. (2012). Coming Across Information Serendipitously: Part 1 – A Process Model. Journal of Documentation. v68 n5

[6] McNee, S. Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K., Rashid, Al M., Konstan, J. and Riedl, J. (2002) On the Recommending of Citations for Research Papers. In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, Louisiana, USA, 16-125

[7] Russ, S. W. (1993) Affect and Creativity: the role of affect and play in the creative process. NJ: Lawrence Erl

[8] Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[9] Sun,X., Sharples,S.,&Kefalidou,G. (2011) User Requirements for an Interactive System to Stimulate Serendipity. Proceedings of Ergonomics & Human Factors conference, 2012, Blackpool, UK

[10] Tintarev, N. and Masthoff, J. (2011) Designing and Evaluating Explanations for Recommender Systems. In Recommender Systems Handbook, chapter 15, Ricci, Rokach, Kior, Shapira and Kantor (Eds), Springer.

[11] Tintarev, N. and Masthoff, J. (2007) Effective Explanations of Recommendations: User-Centered Design. In Proceedings of the 2007 ACM Conference on Recommender Systems, MN, USA, 153-156.

[12] http://max-planck-research-networks.net