

An Ontological Framework for Contextualising Information in Large Hypermedia Systems

Thesis Overview

An overview of the work presented in the PhD dissertation of the same title.

Andrew Bucknell

2007

The popularity and usefulness of the World Wide Web are directly influenced by the massive amount of information available through websites. The ability to access information from around the world using tools such as Web Browsers like Internet Explorer and Firefox, and Search Engines like Google and MSN Live Search has changed the way people perform a range of everyday tasks such as shopping, communicating, research and entertainment. The explosion in the use of these information technologies has lead commentators to label the 21st century the Information Age.

The massive amount of information available on the World-Wide Web is transforming modern society. At the same time the World-Wide Web changes the way people work, communicate and carry out commerce, the same mass of information that makes the World-Wide Web useful also threatens its usefulness. Issues such as information overload, cognitive overload and cognitive disorientation reduce the ability of users to effectively carry out information-seeking behaviours, thus reducing the usefulness of the World-Wide Web.

The technologies of the World-Wide Web have been designed to be open, scalable and extensible. This allows new infrastructure to be built on top of existing protocols and technologies. These technologies allow a user to carry out information-seeking behaviours in a way that has never been possible before. Information-seeking behaviours include browsing and searching the web. This work uses these technologies to build a software framework that can be used to create tools that augment a user's information-seeking behaviour.

Information-Seeking

The World Wide Web is used to find information. Information seeking can be viewed as a form of problem solving. Two basic behaviours have been identified as being used by people seeking information in electronic environments such as the World Wide Web. One kind of behaviour is browsing where the user seeks to satisfy an information need by scanning a collection of information (Marchionini 1997). The second behaviour is searching, where the user utilises a query to convey their information need to a search engine (Broder 2002).

Information seeking behaviour on the World Wide Web involves handling a massive amount of data and data sources. The availability of tools such as browsers and search engines for accessing this information brings with it a set of behaviours that mimic traditional electronic information seeking behaviours, but have their own characteristics. This work draws on the four kinds of information-seeking behaviours shown in Table 1 (Choo, Detlor & Turnbull 2000).

Scanning Mode	Information Need	Information Seeking	Information Use
Undirected Viewing	General areas of interest; specific need to be revealed	"Sweeping" Scan broadly a diversity of sources, taking advantage of what's easily accessible	"Browsing" Serendipitous discovery
Conditioned Viewing	Able to recognise topics of interest	"Discriminating" Browse in pre-selected sources on pre-specified topics of interest	"Learning" Increase knowledge about topics of interest
Informal Search	Able to formulate simple queries	"Satisfying" Search is focused on area or topic, but a good-enough search is satisfactory	"Selecting" Increase knowledge on area within narrow boundaries
Formal Search	Able to specify targets in detail	"Optimising" Systematic gathering of information about an entity, following some method or procedure	"Retrieving" Formal use of information for decision-, policy-making

Table 1 Information Seeking Behaviours on the World Wide Web

These behaviours provide a categorisation of the kinds of behaviours performed when seeking information on the World Wide Web. Such behaviours are augmented when users have useful tools available. This work uses these behaviours as the basis

for applying contextualisation to helping users find information on the World Wide Web.

Context

This work builds on the existing research that demonstrates the significance of context in information seeking behaviours. In particular this work draws on concepts of context developed in information retrieval research (Belkin 2000; Spink 1999). Many of these ideas about context are derived from inter-disciplinary analyses of concepts of context (Dervin 1997). The following points regarding context are of particular interest in this work.

- 1) *Knowledge is partial and temporary.* When perceiving a message or an event, a receiver will only interpret as much of it as is relevant to his or her current interest. Therefore, what a receiver knows about a message or event depends on the context on which it is being received.
- 2) *The knower and the known are inextricably bound.* Each person has a unique collection of *a priori* knowledge, goals and experiences that affect how a message is interpreted. Thus, what a user knows as a result of receiving a message depends on the nature of the user. The result of interpretation of a message is not constant. Because it depends on the knowledge, goals and experiences of the user, the interpretation will vary across users, and will even vary for the same user as their knowledge, goals or experiences change over time.
- 3) *Context is not usefully conceptualised as an independent entity.* It is neither meaningful, nor useful to identify an entity independent of the received message and label it as context. Context is more than just a set of related facts that alter perception of a message. Any such entity is an artificial construct that trivialises the impact of external factors on perception.
- 4) *Context is necessary as a source of meaning.* Communication relies on common experiences and understandings in order that the interpretation of messages by sender and receiver are close enough that the ideas understood by the receiver are those generated by the sender.

In this work, context is treated as information about web-based information. This can either be an element or set of elements from a web page, information derived from the page, or information about the users interaction with page. These elements have all been used in existing approaches to contextualisation. The contribution of this work is to use a knowledge-based approach to modelling these elements of context to support reasoning about a user's intentions to augment their information-seeking behaviours with dynamic and useful user interfaces.

Ontology

The use of formal ontology in information systems is an area of growing interest. Modelling web-based knowledge using ontology is a key project of the Semantic Web initiative, and has led to the development of an XML application called the Web-Ontology Language (OWL) (Heflin 2004). There are five key benefits that arise from representing information using ontology (Guarino 1998). Each of these benefits is discussed in turn below.

Completeness. Because an ontology will be developed to represent context in a domain, rather than just for a single application, the set of entities specified by the contextual ontology provided a useful reference for application developers working within that domain.

Consistency. The elements of an ontology allow us to refer to domain elements in a consistent way, reducing ambiguity and allowing for the construction of thesauri that allow the mapping of vocabularies between groups of specialists operating in the domain.

Formality The rigour and structure imposed on the domain elements being described in a contextual model mean that assumptions about structure are minimalised, and allow for the creation of standardised interfaces that work with the formal structure.

Reusability. Because the contextual models are well defined and can be mapped to different vocabularies, a model created for one application can be used by other applications that are aware of the ontological structure. It is also possible to provide mappings between ontological structures to support this kind of reuse.

Interoperability. Ontologies can be implemented as interfaces rather than as just data elements, so they can be made available as standalone components that can be readily plugged in to other systems or applications. It is also possible to create various tools that utilise ontologies without having to impact on existing tools or systems that are in place.

This work using the Simple Knowledge-Engineering Methodology (SKEM) approach (Noy & McGuinness 2001) to develop an OWL-based model that expresses an ontology of context. This model (Figure 1) describes how context can be represented in a machine readable form that supports reasoning about the elements and the relationships between them.

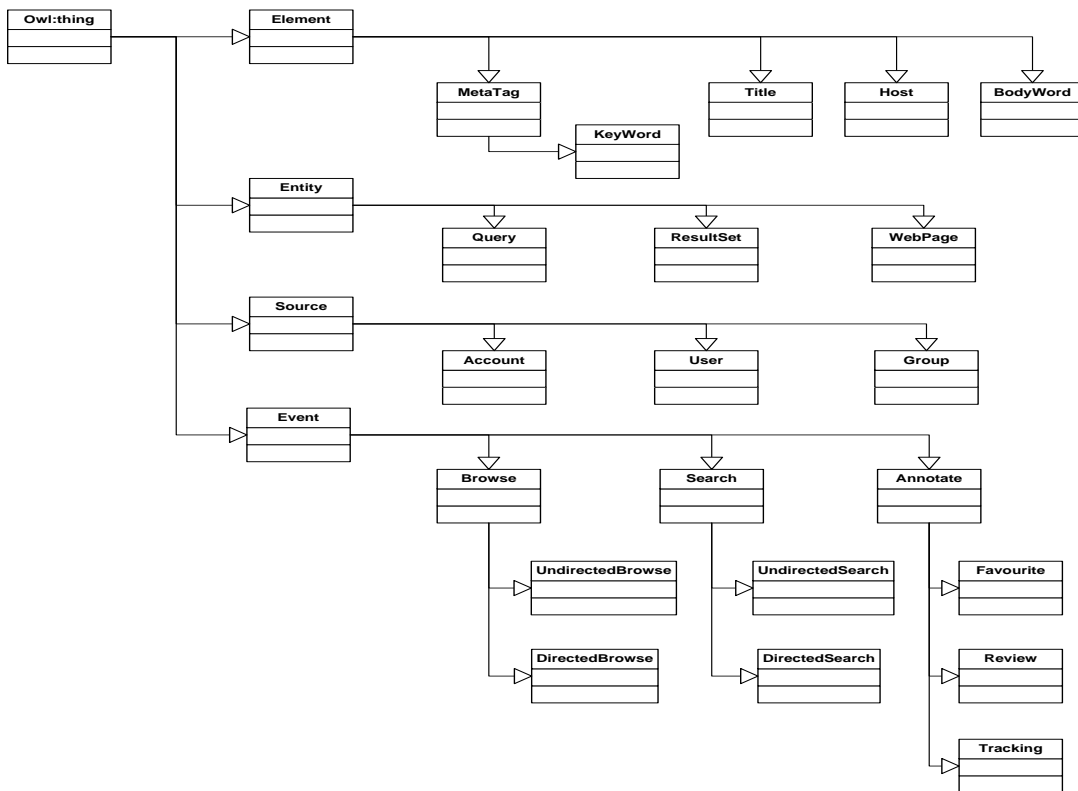


Figure 1 Class view for ontology of context

Interfaces

Interfaces are developed to support scenarios that describe how users seek information using the World-Wide Web. The interfaces make contextual information

available to the user. The scenarios to be supported are based on information retrieval research into how user's browse and search information spaces, as well as on information-seeking behaviours for the World-Wide Web. These scenarios and the user interfaces that support them are developed using Scenario-Based Design (SBD) (Rosson & Carroll 2001). When designing the augmented scenarios that will be supported by a user interface in the domain of users performing information-seeking behaviours, the following principles are considered.

Effective. The scenarios should directly meet the need and goals of a user.

Comprehensible. The scenarios should be clear. Their application to the problem domain and their usefulness should be readily apparent.

Satisfying. The scenarios should improve the experience the user has when using the system to carry out tasks.

The user interface scenarios are derived from the information-seeking behaviours for the web. Some of the behaviours have been relabeled to be more inline with information-retrieval research. The behaviours are described in terms of user behaviours and information sources in Table 2.

Directed Search	In directed search, a specific set of data is being searched to satisfy a specific information need.
Directed Browsing	The user browses a specific information store to locate information that satisfies a specific information need.
Undirected Browsing	The user browses unspecified sources for an unspecified information need.
Undirected Searching	The user searches an unspecified set of information sources for a specified information need.

Table 2 Information-seeking behaviours.

A set of scenarios can be identified that are representative of these behaviours. There are many possible scenarios that could be included here. For this work, scenarios that are effective, comprehensible and satisfying were selected. Two scenarios were chosen for each behaviour. In practice many more scenarios are

possible. The model and framework developed in this work has been designed so that further scenarios can be supported. The scenarios identified in Table 3 have been selected as candidates for our proof of concept implementation.

Behaviour	Scenario
Undirected Browsing	View Page Info
	View Related Pages
Directed Browsing	View Hotlist
	View Dynamic Favourites
Undirected Search	Browse Result Set
	Search Result Set
Directed Search	Browse Context Store
	Search Context Store

Table 3 Information-seeking behaviour scenarios.

Framework

This work develops a software framework that supports contextualisation. Contextualisation is described in this work as a process consisting of three asynchronous activities. Contextualisation involves Identifying contextual information, Collecting contextual information, and Using contextual information. This framework can be reused to develop new components for identifying, collecting and using contextual information. The framework provides us with a reusable and extensible architecture that allows new interfaces that support further scenarios of information-seeking behaviour.

Identifying. Extract contextual information from documents the user views, monitor the user's interaction and navigation patterns to extract further contextual information.

Collecting. Store the contextual information in a knowledge store that supports reasoning about the contextual information.

Using. Make contextual information available to the user.

This collection of activities has been labeled the ICU paradigm in this work. In order to apply this paradigm to developing tools that help people find information on the World-Wide Web, a software framework that describes these activities in terms of software components has been developed. This framework is referred to as the ICU framework (Figure 2 Top-level ICU Framework Architecture).

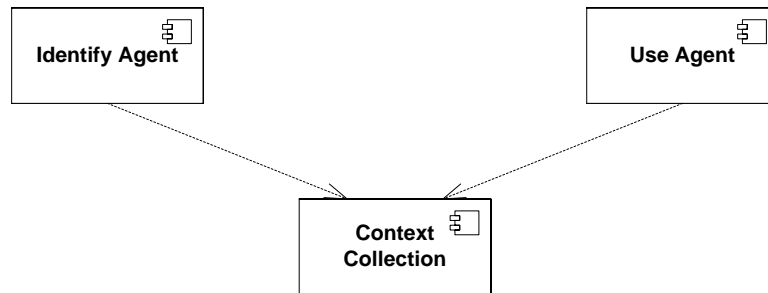


Figure 2 Top-level ICU Framework Architecture

ISeeYou

ISeeYou is a tool developed using the ICU Framework that is used for contextualising web-based information. ISeeYou consists of three software components. Each one is an instantiation of one of the components of the ICU framework shown in Figure N. The first component is an implementation of a Collect component, collectApp. The collectApp component is implemented in Java. It publishes a set of Web Services using SOAP. These interfaces are used by distributed components to update and query the collection of contextual information.

The second component, identifyApp, is an implementation of an Identify component. It is implemented as a Browser Helper Object using C++ and ATL. It is loaded by Internet Explorer and monitors the user behaviour using the DWebBrowserEvents2 dispatch interface. Contextual information recognised by the identifyApp is encoded in a SOAP message and sent to the collectApp. In this implementation it is assumed that the collectApp is running on the same machine.

The third component, useApp, is an implementation of a Use component. It is implemented as an Explorer Band for Internet Explorer using C++ and ATL. The user interface is provided using HTML. Contextual information retrieved from the collectApp is encoded using XML. The XML is run through an XSLT script to generate

HTML that is presented to the user. This allows the user interface to be easily modified or customised as data elements are changed or added, and as new user interface (UI) techniques are identified as being useful to the user. The contextual information is retrieved by sending SOAP-based queries to the collectApp when changes in the content being displayed are detected, or when the user selects a new option in the Explorer Band.

Evaluation

The usefulness of ISeeYou for information-seeking behaviours on the World-Wide Web was evaluated using the Technology Acceptance Model (Davis 1989). TAM has been repeatedly shown to be a reliable instrument for predicting the adoption of a new technology for a particular task (King & He 2006; Legris, Ingham & Collette 2003). The factors measured by TAM are Perceived Usefulness (PU) and Perceived Ease Of Use (PEOU). Perceived usefulness is defined by Davis as 'the degree to which a person believes that using a particular system would enhance his or her job.' It is a measure of the extent to which a user believes a system would help them with the task they are trying to achieve. Perceived ease-of-use is defined by Davis to be 'the degree to which a person believes that using a particular system would be free of effort.' It is a measure of how easy a user finds a system to use for the task they are carrying out. These constructs are measured by a survey that uses a set of established questions to collect user's responses. The questions used to measure PU are shown in Table 4. The questions used to measure PEOU are shown in Table 5. Responses are measured using a 7 item Likert scale ranging from Extremely Likely through to Extremely Unlikely.

Perceived Usefulness
PU1 Using I See You helps me find information on the web more quickly.
PU2 Using I See You improves my performance when finding information on the web.
PU3 Using I See You can increase my productivity when finding information on the web
PU4 Using I See You can enhance my effectiveness when finding information on the web

Table 4 Scale items of the usefulness determinant

Perceived Ease-of-Use
PEOU1 Learning to use I See You is easy for me.
PEOU2 I can use the I See You in a manner that helps me find information.
PEOU3 My interaction with I See You is clear and understandable.
PEOU4 In general, I find I See You easy to use.

Table 5 Scale items of the ease of use determinant

The TAM Evaluation was performed using students from three Masters level subjects at the University of Technology, Sydney. These subjects were chosen because it is expected that students have a reasonable level of proficiency using the web to find information, and are also expected to use the web for study-related activities during the subject. Participation in the evaluation was voluntary and anonymous in accordance with the guidelines of the UTS Human Research Ethics Committee.

Participants were asked to download and install ISeeYou. They then used ISeeYou while they engaged in their regular information-seeking behaviours, whether study related or otherwise. Participants were then prompted to fill out the survey instrument, which was hosted on the Institute for Interactive Media and Learning's Survey Manager application. The results were collated and analysed using a Factor Analysis and a Reliability Analysis. The scale items were found to measure the expected two factors, and the Cronbach's alpha calculation for the factors indicated that the measurements were reliable. The medians for the factors are shown in Table 6.

	Median
PU	3
PEOU	3

Table 6 PU and PEOU for ISeeYou

The median Perceived Usefulness of ISeeYou is **3**. From this it can be concluded that it is **likely** that ISeeYou is perceived to be useful. The median Perceived Ease of Use of ISeeYou is **3**. From this it can be concluded that it is **likely** that ISeeYou is perceived to be easy to use. These results indicate that a tool built using the ICU framework is perceived to be useful for people performing information-seeking behaviours on the World-Wide Web. This result supports further research into the

use of the ICU framework as an approach to developing tools for Personal Information Management and tools for contextualising web-based information.

Conclusion

The framework developed in this work, based on an ontology of context, can be used to create tools that users find useful when carrying out information-seeking behaviours on the World-Wide Web. Several interesting avenues of research can be pursued using this framework, including developing new interfaces to support information-seeking behaviours, extending the ontology of context to support a richer set of elements and relationships, and refining the evaluation approach to provide more fine-grained information about the usefulness of interfaces developed using the ICU framework.

In addition to potential research applications, this work has several commercial applications, focused around augmenting user ability to find information on the World-Wide Web. Providing a personal context store that supports reasoning about a user's history would allow browsers to support interfaces that adapt to a user's interests and make the browsing experience more useful. The context store could also be used to provide client side filtering of search results, using knowledge-based techniques to highlight search results more likely to be of interest to the user. On a broader scale, the context store serves as a user's Database of Intentions as described by John Battelle (Battelle 2005). A user's Database of Intentions is the aggregation of their search queries and the link anchors they clicked. Having a user's Database of Intentions stored in a structure that supports reasoning allows many inferences about a user's interests to be made. By sharing elements of a context store between users from a community, a community-based Database of Intentions could be constructed, allowing inferences about the interests and intentions of groups of users to be made. These inferences could be used to provide community based link histories, to filter search results based on community interests, and to provide information targeted at the community such as advertising. The range of possible applications of this work is broad and of much potential benefit to information producers and consumers alike.

References

- Battelle, J. 2005, *The search: how Google and its rivals rewrote the rules of business and transformed our culture*, Portfolio.
- Belkin, N.J. 2000, 'Helping people find what they don't know', *Communications of the ACM*, vol. 43, no. 8, pp. 58-61.
- Broder, A. 2002, 'A taxonomy of web search', *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3-10.
- Choo, C.W., Detlor, B. & Turnbull, D. 2000, 'Information Seeking on the Web: An Integrated Model of Browsing and Searching', *First Monday*, vol. 5, no. 2, pp. 67-78.
- Davis, F.D. 1989, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS Quarterly*, vol. 13, no. 3, pp. 319-340.
- Dervin, B. 1997, 'Given a context by any other name: methodological tools for taming the unruly beast', *Proceedings of an international conference on Information seeking in context table of contents*, pp. 13-38.
- Guarino, N. 1998, *Formal ontology in information systems*, IOS Press.
- Heflin, J. 2004, *OWL Web Ontology Language Reference*, 2006 <<http://www.w3.org/TR/owl-ref/>>.
- King, W.R. & He, J. 2006, 'A Meta-Analysis of the Technology Acceptance Model', *Information and Management*, vol. 43 (2006), pp. 740-755.
- Legris, P., Ingham, J. & Colletette, P. 2003, 'Why do people use information technology? A critical review of the technology acceptance model', *Information and Management*, vol. 40, no. 3, pp. 191-204.
- Marchionini, G. 1997, *Information Seeking in Electronic Environments*, Cambridge University Press.
- Noy, N.F. & McGuinness, D.L. 2001, 'Ontology Development 101: A Guide to Creating Your First Ontology', *Knowledge Systems Laboratory, March (2001)*.
- Rosson, M.B. & Carroll, J.M. 2001, *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*, Morgan Kaufmann.
- Spink, A. 1999, 'Towards a theoretical framework for information retrieval in an information seeking context', *Exploring the contexts of information behaviour table of contents*, pp. 21-34.