Towards Semantic Web Support for Writing Project Proposals

Timothy Miles-Board
Intelligence, Agents, Multimedia Group
University of Southampton, UK
tmb@ecs.soton.ac.uk

Leslie Carr
Intelligence, Agents, Multimedia Group
University of Southampton, UK
lac@ecs.soton.ac.uk

Arouna Woukeu
Intelligence, Agents, Multimedia Group
University of Southampton, UK
aw1@ecs.soton.ac.uk

Gary Wills
Intelligence, Agents, Multimedia Group
University of Southampton, UK
gbw@ecs.soton.ac.uk

Wendy Hall
Intelligence, Agents, Multimedia Group
University of Southampton, UK
wh@ecs.soton.ac.uk

ABSTRACT

Writing technical project proposals is a task that needs convenient access to a wide range of information. The Web provides some support for this activity in the form of search engines (to discover relevant material) and links (to make reference to supporting evidence). This paper establishes some requirements for supporting proposal authoring and describes how they can be satisfied by a Semantic Web environment.

1. INTRODUCTION

The aim of the Semantic Web is to make the present web more machine-processable, in order to allow intelligent agents to better retrieve and manipulate pertinent information. However, the end client is usually a human, rather than machine, intelligent agent. Many Semantic Web applications aimed at humans have been concerned with search, *i.e.* the use of ontological descriptions to improve the location and retrieval of relevant information, often through a semantic portal [11, 2, 19]. Other applications have focused on annotation - the ability to improve the display of information about known concepts which appear in a text and which feature in an ontology [6, 4].

Both of these kinds of applications address readers of the Web, but Semantic Web technologies can equally well address the needs of writers of the Web. A writer, especially of technical or business material, needs to synthesise information from various sources in order to create new documents.

In the context of the Web, the writer can take advantage of a search engine to find information that is required and should then link the useful resources to the new document. Various hypertext solutions assist the user in these tasks [21, 18]. However, the writer still has to undertake a substantial reading task in order to write - firstly locating useful documents and then reading and understanding the contents before extracting the required information.

WWW2005, May 10–14, 2005, Chiba, Japan.

In the context of the Semantic Web, a semantic model of the contents of the documents may well be available for use by the author if the document's contents have already been marked up against a suitable ontology. Other Semantic Web processes and services may have harvested and cached that semantic information in order to offer services to just such an author. Consequently the tasks that the author needs to undertake in the Web context (search, read, understand, extract) may potentially be streamlined in the context of the Semantic Web (directed lookup).

This paper describes the work of the WiCK project[1], whose aim is to investigate the potential for authorship support that an existing Semantic Web knowledge store can afford. So as to make such an investigation practical, we have focused on a specific kind of writing task, *i.e.* producing a project proposal. The outputs of this task fall into a well-defined genre, whose usage (by funding bodies' grant committees) means that the type of material that the author needs to produce (and hence the type of material which the author needs to use) is specified and constrained quite precisely.

Even so, writing project proposals is a very involved task with a large number of information requirements, from costings and work-plans through objectives and related work analyses. Complex arguments must be formulated to support a particular course of action and to justify expenditure. For each argument, evidence must be presented that is attested by work that appears in the literature or by other projects that have been previously funded.

The work that this paper describes does not claim to relieve the author of the effort required to construct these arguments, but it does attempt to address the issue of what facilities should be provided and how they could be delivered to the author in order to provide demonstrable assistance in the effort of creating new (and convincing) proposal documents. In order to achieve this, a new writing tool is proposed which works in conjunction with an existing (and proven) knowledge store.

[1] `http://wick.ecs.soton.ac.uk/`

2. WRITING PROJECT PROPOSALS

The task of writing a funding proposal is common in industrial and commercial environments; here, we consider a funding proposal for a research project in an academic environment. The proposal is directed at the UK's Engineering and Physical Sciences Research Council (EPSRC), which has a well-defined procedure for submitting, reviewing, and selecting proposals for funding, and provides a standard form[2] (the Je-SRP1) and a comprehensive guidance document[3] on how to fill out the form, create the supplementary documentation, and submit it for consideration. The University of Southampton is well positioned for investigating this particular scenario, as it is the 4th highest beneficiary of EPSRC funds in the UK[7].

The Je-SRP1 form itself serves as an administrative summary of the research proposal, collecting together the relevant information about the hosting organisation, project investigators, project partners (for joint proposals), referees, staff (including visiting researchers), and travel and equipment costs. The 'meat' of the proposal is contained in the supplementary document — the Case for Support — the composition of which is tightly defined in the guidance notes. The rules for the Case define the formatting (constraints on page length, font sizes *etc.*), the information content, and the structure of parts and sections where each of these pieces of information should be placed.

The Case for Support has two parts. Part 1 requires the author to "provide a summary of the results and conclusions of recent work in the technological scientific area which is covered by the research proposal", with reference to both EPSRC-funded and non-EPSRC funded work, and to outline the specific expertise available for research at the organisations involved in the proposal. In other words, the author needs to *report from the cutting edge*, placing the proposed work amongst other efforts on this research frontier and give evidence that the project partners have the necessary expertise to push back the boundaries of this frontier further.

Part 2 of the Case for Support has two main themes; topic background and work-plan. The former requires a fuller description of the topic of research and its academic and industrial context, instructing the author to "demonstrate a knowledge and understanding of past and current work in the subject area both in the UK and abroad", that is to *provide a 'historical' perspective* to the proposed work. The latter theme of Part 2 is the *proposed work-plan* for the project, for which the author must describe the project programme and methodology (including overall aims and objectives), identify the potential impact of the work and its relevance to beneficiaries, indicate proposed dissemination and technology transfer routes, and justify the resources (staff, equipment, travel, services etc.) requested for carrying out the research.

In summary, the main writing tasks in completing a funding proposal are:

- Filling in the form.
- Reporting from the cutting edge.
- Providing an historical perspective.
- Proposing a work-plan.

[2] `http://www.epsrc.ac.uk/website/commonpages/downloads.aspx?CID=4482`

[3] `http://www.epsrc.ac.uk/website/commonpages/downloads.aspx?CID=8621`

2.1 Investigating Current Practice

To begin to put together a picture of current practice in writing EPSRC project proposals, we carried out a focused investigation within our department. A group of 10 participants were interviewed, each with varying degrees of experience of proposal writing, but each belonging to the School of Electronics and Computer Science (ECS) — our participant pool was therefore drawn from a highly computer literate environment. We split the pool into two groups of five participants, novices and experts, reflecting their relative proposal writing experience.

In the context of the WiCK project, we were specifically interested in the kinds of informational resources that the authors engage with during the main information-intensive tasks identified in the previous section: filling in the Je-SRP1 form, reporting from the cutting edge, providing an historical perspective, and proposing a work-plan.

2.1.1 Information Sources

To provide a starting point for discussion we seeded a list of 14 different information sources which we thought could potentially be useful to an author preparing a funding bid. As it turned out, this list was broad enough in scope to attract only a few additions by participants during the investigation (see below). Table 1 summarises these sources, which we have classified into four categories: *personal/interpersonal. institutional*, *funding body*, and *World-Wide Web.*

Within the *personal/interpersonal* group, we anticipated that the author's own recollection and experience could provide information needed for the various tasks. For the purposes of this particular investigation, we broadly distinguished *recall* as recovering required information by mental effort (*e.g.* remembering John's phone number), and *experience* as the recollection of a process by which the required information has been obtained in the past (last time I asked Mary from Finance for John's salary details). In both cases, actual grey cells could be augmented by notes (*e.g.* academic logbooks). Additionally, authors could ask their colleagues for information, for example through formal discussion groups, research mailing lists, or informal chats over coffee.

The *institutional* group represents those sources of information available to the author within the department and/or institution at which they are based, in this case the Electronics and Computer Science department. To this end, authors may direct enquiries to the human resources department, or extract information from the ECS intranet (projects, people, publications), academic CVs for members of the department, or previous EPSRC project proposals completed by the author or by others in the department.

The EPSRC itself provides a number of resources which could potentially prove useful to authors putting together a funding bid. The EPSRC project pages provide information about previously and currently funded projects. Recently introduced by the Research Councils UK group, the Je-S service assists applicants in completing the Je-SRP1 form by automatically filling in required information where available (provided that this information has previously been registered by the named individuals and/or organisations). Lastly, as we have already seen, the EPSRC guidance notes provide a wealth of information and advice to applicants.

Personal & Interpersonal	Recall
	Experience
	Colleague
Institutional	Personnel
	ECS Website
	Academic CVs
	Previous Proposals
Funding Body (EPSRC)	Projects Pages
	Je-S
	Guidance Notes
World-Wide Web	Search Engine
	Lit. Services
	Digital Library
	Homepages

Table 1: List of information sources, with groupings, used as starting point for discussion.

Finally, the *World-Wide Web* group represents those information sources outside of the institutional and funders' boundaries. Potentially valuable sources of information include literature-based services such as CiteSeer's citation indexing [14], the electronic materials archive by digital libraries (*e.g.* ACM Portal, IEE INSPEC), and homepages of people, projects, research groups, communities etc. Such pages, as well those containing other required information, are likely to be discovered through the use of a Web search engine (*e.g.* Google).

2.1.2 Method

The investigation took the form of an informal directed interview in which participants were initially given a chance to re-familiarise themselves with the EPSRC guidelines, particularly in reference to the four main writing tasks that we wanted to focus on. The participants were then asked to identify from our list, in the context of each of the tasks, the information sources which they had used in the past when preparing a proposal, and to indicate those sources which they considered to have been most important. Participants were also encouraged during the course of the interview to add any information sources to the list that they felt had been overlooked. Some participants also had paper or electronic copies of their proposals to hand, and referred to them. Participants were asked to jot down anecdotal notes that occurred to them during the interview, relating, for example, to the type of information extracted from a source and the reason for extracting it.

2.2 Investigation Results

The numerical results of the investigation are presented in Figure 1. Perhaps the most striking feature in these graphs is the diverse range of information sources our participants had consulted during the proposal writing process — each information source on our list was ticked at least once; most were ticked many more times. In fact, four participants each extended our list with an additional information source:

Tech. Support Kit-related queries and costings directed at the department's technical support and equipment purchasing staff assisted expert in the form-filling task.

External Services A cost calculating service provided by an external website was used by one novice participant in the form-filling task.

Final Reports In addition to it's proposal documents, the *final report* of a past project proved very useful for one novice participant in the historical perspective and cutting edge tasks.

Call for Proposals Proposals answering a specific call issued by the funding body are subject to a stricter set of requirements (for example, a recent call required bid to demonstrate how their proposed research fitted into an architectural framework outlined by the funding body) and so the *call for proposals* was an important source of information for one expert participant in the work-plan task.

The figures show that experts are more likely to consult a wider range of information sources more often during proposal writing than novices. However, the highest scoring group of sources in each of the four tasks were corroborated by both novices and experts: institutional sources score highest in the form-filling task; in the cutting edge task the WWW sources score highest; in the historical perspective task both personal and WWW groups feature highly, with personal sources just having the edge; finally, personal sources score highest in the work-plan task by a wide margin.

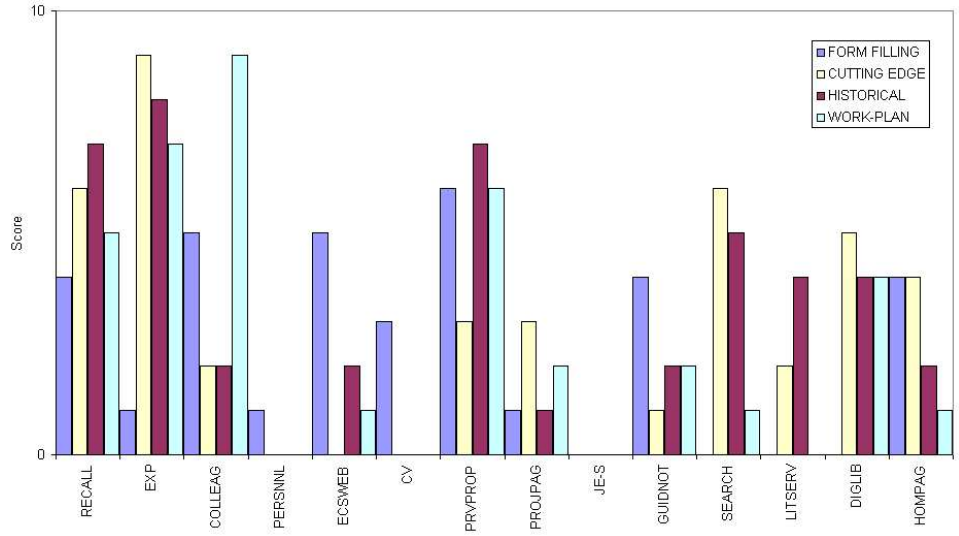
In terms of individual sources, previous project proposals were rated highly throughout the four tasks, second only to recall and experience overall; this information source also attracted many comments from participants. One such comment seems to typify this trend: "look for related material that could be used as a basis for the first draft"; predictably therefore, the lowest score for this source was reported in the cutting-edge task — what was on the cutting edge six months ago has since become history so information for this task will have to come from more up to date sources!

Participants reported different levels of information reuse from previous proposals, ranging from looking at "similar topic proposals to get an idea of current trends" and identifying "key information to go in the *Programme and Methodology section*", through reflecting on "style, good practice, phraseology" as used for "successful vs. unsuccessful proposals", to cut and pasting "boilerplate text" and figures ("what can I reasonably ask for a workstation? — grab figures out of recently accepted proposal"). One novice even ranked previous project proposals as being a particularly important source, even though none were actually available to him at the time that he had produced his only EPSRC proposal to date.

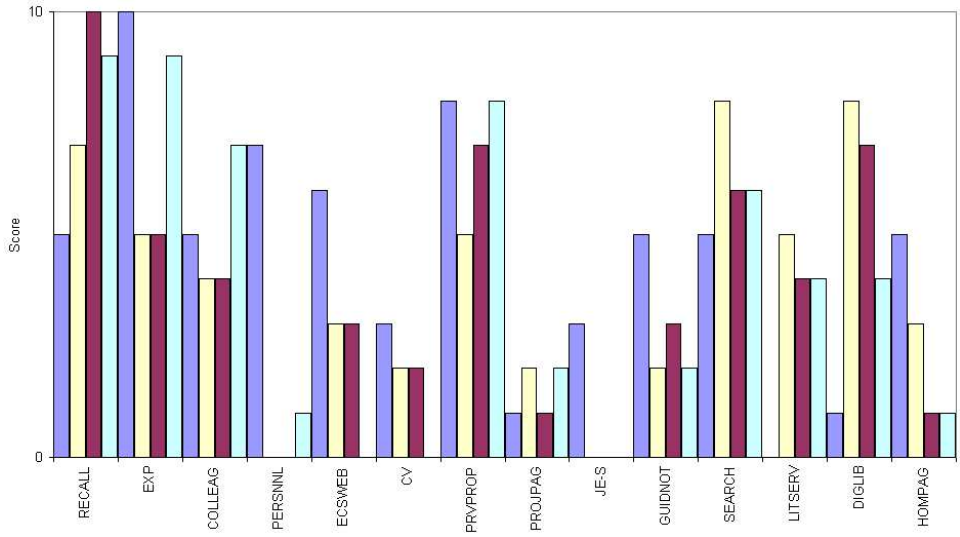
One expert who highlighted the importance of extracting information from previous "successful proposals (not only my own!)", lamented that "the people with the most experience of writing project proposals are always the busiest" (and therefore less likely to be able to find time to answer colleagues' questions about their own proposals). To help himself and his colleagues in future, he suggested an intranet-based "pool of previous proposals from the research group". This notion was also independently put forward by another (expert) participant, who drew parallels with corporate memory systems.

2.2.1 Filling in the form

In this task, experts relied more on experience (presumably they had more of it —"what I did last time"), and



a. Novice Scores.



b. Expert Scores.

Figure 1: Graphs showing scores given to each information source by both novice and expert participants (each source has a maximum score of 10).

conducted many more personnel enquiries (this could be accounted for by the fact that the staff costing duties usually fall to the senior member of the project team *i.e.* the 'expert'). Whereas novices did not use search engines at all in this task, examples of search engine use from experts included "checking that acronym for proposed project is not in use" and "looking up project partner information". It is interesting to note that only experts reported using the Je-S electronic application service, which likely reflects the relative newness of this resource.

2.2.2 *Reporting from the Cutting Edge*

Novices relied more on experience in this task; rather than place as much emphasis on this source, the experts consulted a wider range of resources more often. Perhaps unsurprisingly, the WWW group of sources were scored more highly in this task than any other — participants' comments revealed that using these sources was the best way to uncover the "major new initiatives" in project's research area. A number of recall- and experiential-based strategies ("using stuff you know to find out stuff you don't know") for using these resources came to light, including "looking at most recent work of relevant authors and research groups", following up "emails from friends, things heard in discussions at conferences", and "checking homepages for new material from known experts". One expert provided further insight, noting that his personal recollection and experience acted as "drivers to solve the problem", leading to "extensive literature searches" using the WWW sources in an iterative fashion against "a set of rules based on what is in the proposal".

2.2.3 *Providing an historical perspective*

One expert ranked digital libraries as "most important

for demonstrating historical context through references", a comment which echoes many of the other insights recorded in this task. For example, one novice noted that she looked at ECS intranet pages to assess "relevant expertise" from within her research group, used a digital library to "follow up references" from these pages, falling back to a search engine to "track down electronic version of papers not in the digital library". Other participants adopted similar approaches: using WWW sources to flesh out the "details of known", "fill out a bibliography", or as an aid to "justify the starting point for discussion". An (expert) participant reported that during this process he also took into consideration "who will review, who will be on the panel?" when evaluating which material should and should not be included in the proposal. One novice writing a proposal for work in a different research area to his own found himself "reinventing my past" through targeted web searches for combinations of keywords to find relevant papers.

2.2.4 Proposing a work-plan

Experts placed a greater emphasis on recall and experience in this task (presumably having a greater wealth of experience to draw upon than novices). The experts also scored the WWW group of information sources much higher than the novices, although still less than in the cutting edge and historical perspective tasks. Comments by participants reveal some of the functions carried out by the WWW sources: "check that new terminology I make up does not have conflicting meaning in others areas", "use search engine to explore beneficiaries and dissemination routes", visit homepages of industrial sponsors to "cover their interests", cite relevant work on which methodology is based — "as done by *X* evaluation". Both novices and experts also ranked their colleagues higher than in any other task. One novice noted that she had talked over her work-plan with a senior colleague to "get an idea of the feasibility of the project in terms of goals and timing. An expert, on the other hand, might seek to find out from a colleague the "latest innovations in dissemination".

2.3 Requirements for Supporting Proposal Writing

In order to better inform computer-based support for the proposal writing process, we extracted a set of requirements that such systems should meet. Perhaps the most arresting result from this investigation is the diverse range of information sources participants interacted with during the proposal writing process — our small study revealed regular use of at least 16 different information sources during the process. It is also clear that there is a great deal of complex interaction taking place between the author and these sources, for example the (novice) participant that started her information retrieval task by consulting intranet pages to assess relevant expertise within her research group, moving to a digital library to follow up useful looking paper references from these pages, and then falling back to a search engine to try and track down those papers that could not be accessed. Computer-based support for the proposal writing task should therefore be able to handle multiple different information sources.

In order to carry out the proposal writing task, authors must necessarily contend with, and ultimately overcome, a high cognitive overhead and 'forced divided attention' [21], not only dealing with multiple different information sources during the task but also having to switch between the different interaction mode of each source, *e.g.* picking up the phone to call a colleague, firing off an email, or opening a Web browser and surfing to any number of different locations. The fact that in our investigation experts consulted more information sources more often could be interpreted as authors becoming more attuned to this process as they become more experienced at it, and that such skills must therefore be learnt by novices. Computer-based support for proposal writing should therefore also facilitate better management of these sources, for example, reducing 'forced divided attention' by providing information at the author's fingertips in their chosen editing environment. This could also include the leveraging of expert strategies to help novices make their writing more effective — examples observed in our investigation included considering who will be reviewing the proposal (are panel members interests and/or research efforts covered?) and using the WWW group of sources to justify the proposed methodology where appropriate.

It is also clear from the investigation that there is a requirement to support different information sources and usage strategies in different tasks: institutional sources scored highest in the form-filling task, WWW sources in the cutting edge task, personal and WWW sources in the historical perspective task, personal resources in the work-plan task. A further requirement for computer-based support is therefore the ability to provide informational services which most closely match the task or subtask that the author is currently engaged in.

A number of opportunities for information reuse are apparent in the proposal writing task, the most obvious being the wide range of strategies reported by participants in extracting useful information from previous project proposals. A final requirement for computer-based assistance is therefore that the ongoing evolution of its information sources should be supported. For example the output of one author's proposal writing task — the finished funding bid, in the EPSRC case the completed Je-SRP1 form and accompanying Case for Support — itself becomes a part of the array of informational sources that the system subsequently makes available to other proposal authors.

In summary, therefore, we have identified the following requirements arising from our investigation of current practice in project proposal writing:

1. Support a large, diverse range of information sources.
2. Manage these sources effectively.
3. Support the differentiated informational requirements that arise from different tasks.
4. Facilitate the reuse of completed proposals.

3. REQUIREMENTS VS. EXISTING APPROACHES

From our investigative work, we have established that the defining characteristic of this problem domain is the synthesis of information from a diverse range of sources. One of the fundamental aims of the Semantic Web [3] is to provide homogeneous access to heterogeneous data: formally specified ontologies codify agreed meanings across diverse

information sources, unifying them through a common language model. In such an environment, computational agents can unambiguously determine the meaning of a resource and its relationship to other (meaningful) resources, thus making the Web an environment in which software agents and humans can make better (reasoned) use of the available resources. The key components of the Semantic Web are therefore (a) agreed models (ontologies) of the objects and relationships contained in the information sources (b) formally specified ontology languages for unambiguously codifying these agreed models and (c) an annotation mechanism for identifying (parts of) Web documents and other sources with concepts from relevant ontologies. Semantic Web technology therefore makes a suitable base on which to build a computer-based environment to support proposal writing by specifically enabling the crucial requirement of proposal writing — access to diverse information sources.

A number of efforts in this area have contributed to our understanding of how Semantic Web technology can be used to assist authors in carrying out different writing tasks. ARIA [17], for example, supports email or web page authoring based on a semantically annotated photo database. By continuously monitoring the text typed by the author against a domain ontology, ARIA recommends photos from the database that seem appropriate to illustrate the various facets of the unfolding narrative. CREAM [9] helps the writer produce the text itself, by dragging and dropping knowledge fragments from an ontology browser into a text editor — for example a dropped slot inserts a text rendering of the slot value (with a link back to the source).

The potential research and commercial benefits of bringing these knowledge-aware processes into the office arena have not gone unnoticed. Microsoft Word, for example, is the most commonly adopted product for authoring text documents [23]; authors can therefore adopt new knowledge-aware extensions without learning a new production environment and without sacrificing familiar features [24]. SemanticWord [23], a Microsoft Word-based environment, adds several toolbars to the standard interface which support the creation of semantic annotations in documents and templates according to selected ontologies (local or imported from the Semantic Web). Annotations are "carried over" in text cut/copy and paste operations, facilitating a level of knowledge reuse between documents. SemanticWord also offers a more proactive annotation feature which the author experiences through the Microsoft Smart Tags interface: as the author types the text content of the document, it is processed by an information extraction component which relates instances and values appearing in the text to ontology instances and types, visually highlighting the matched text in the document. Through the Smart Tags "action" menu, the author can examine the highlighted entities and convert them into semantic annotations.

Although provoking a range of reactions upon its release [12], Smart Tag technology has also been adopted by other office-based initiatives, including SemTalk [8] and OntoOffice [20]. As with SemanticWord, recognised concepts and instances are highlighted with Smart Tags. However, the kinds of action offered differs between systems: in SemTalk, for example, the author can access and edit the underlying ontological model; in OntoOffice, a search for context-relevant documents can be initiated.

3.1 Summary of Shortcomings

Table 2 attempts to match the work described above to the requirements that we have identified for computer-based support for proposal writing: it is evident from this table that there is no single system which meets every requirement.

Since each approach is ontology-based, with instances potentially harvested from a variety of sources to populate a range of different ontologies for use within the system, all the approaches could potentially meet our first requirement to support access to a large range of information sources. In terms of management of these multiple sources however, we can observe two main interaction approaches — CREAM's split screen interface allows authors to add information to their document by dragging and dropping from ontology browser to editor (and annotate their document by dragging information in the reverse direction, from selected text in the editor into slots in the ontology browser); the other approaches use a suggestion-based mechanism based on the recognition of concepts/instances as the user enters keystrokes into the document. In the former case, the onus is on the author to initiate each interaction; the latter mechanism offers a more proactive approach in which suggested information or actions are presented as and when a suitable context is detected (*e.g.* the author types a recognised identifier)[4]. Of the suggestion-based approaches, ARIA and OntoOffice seem most promising, since the recognition is used as a basis to offer further (in context) information (*e.g.* suggested photos) and services (*e.g.* search document repository), that is to *assist the writing task*, rather than simply to "minimise the burden of annotation" [24] (*e.g.* convert recognised text into a semantic annotation) or to enable validation and consistency checks on the document.

Of all the approaches, it seems that only SemanticWord could potentially provide support for the different informational requirements of different tasks by pre-preparing templates for each task. However, the scope of each template extends only to specifying knowledge placeholders which will be populated by authors creating documents based on that template. All of the approaches except ARIA (which merely produces an illustrated text) produce a knowledge-rich document as output, from which ontological instances can potentially be harvested for reuse by other authors.

4. WICKOFFICE

In response to the shortcomings of existing work described in the previous section, we have focused our efforts on a developing a bespoke office-based solution: WiCKOffice. Figure 2 illustrates the features of this Semantic Web environment.

In order to properly model the multitude of different information sources used by authors during the proposal writing task, and hence be able to deploy it usefully in a computational environment, our scenario requires a number of ontologies. To understand and model *what is being written about*, we define a *research ontology* to describe the stake-

[4]Of course, one of the drawbacks of this approach occurs when the knowledge base is incomplete — in the worst case, no terms are recognised and therefore no assisted writing services offered. SemTalk uses WordNet as an external glossary to increase opportunities for recognition; SemanticWord uses NLP techniques to try and extract new instances of existing concepts from sentences.

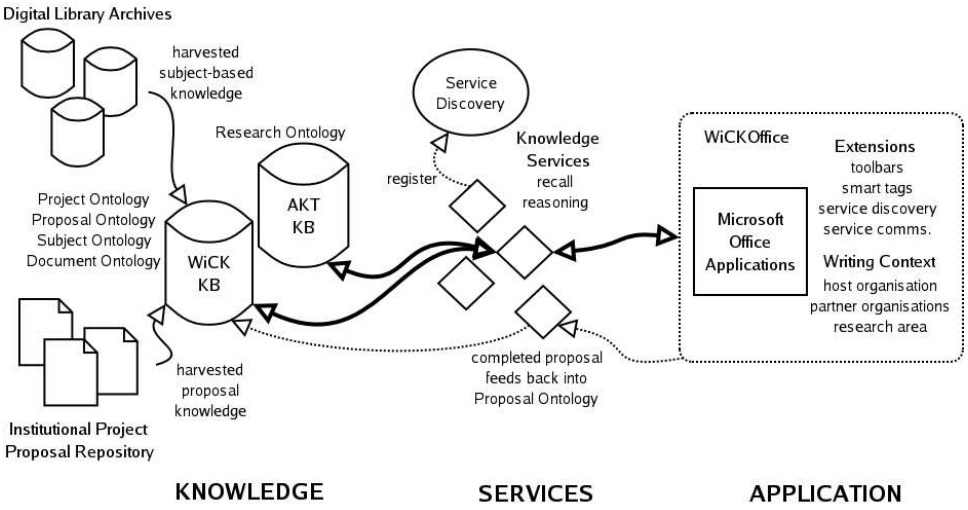


Figure 2: The generalised WiCKOffice knowledge writing environment.

	Requirements			
	Range of Sources	Manage Sources	Diff. Tasks	Reuse Info.
ARIA	●	●	·	·
CREAM	●	●	·	●
SemanticWord	●	○	○	●
SemTalk	●	○	·	●
OntoOffice	●	●	·	●

Key: ● Strength ○ ↔ · Weakness

Table 2: Matching existing work to requirements for computer-based support for proposal writing.

holders and activities who participate in research — the researchers, their publications, research interests, conferences and journals, and a *subject ontology* to describe the area in which we wish to conduct research, the problems that we wish to address and the methods, systems and approaches which have been described in the literature.

The 'design specification' for the proposal writing task itself — *what needs to be written* — is then modelled by a *document ontology* to make explicit the semantic structure of the proposal documents — the pages, sections, paragraphs, forms, and fields. The *type* of information that the author must enter into each part of this structure, is then captured by a *project ontology* — the activity of undertaking work; the ideas of work package, budget, personnel, milestones *etc.* — and a *proposal ontology* — describing the objectives, beneficiaries, funding call, and programme of activity for the project.

Knowledge is managed by two knowledge-bases, both based on the AKT 3Store platform [10]. The AKT knowledge-base models the UK Higher Education computer science community [15] (expressed using the AKT Reference Ontology[1]), harvesting knowledge from multiple sources including home pages and departmental web sites and currently storing in the order of 10 million triples. This knowledge store is already utilised by a number of applications, perhaps most notably CS AKTive Space [22].

A separate WiCK knowledge-base hosts the additional ontologies. Instances for the proposal ontology are acquired from previous EPSRC project proposals; we envision Semantic Web agents trawling digital library archives and automatically constructing and populating the subject ontology. WiCK extensions to the Microsoft Office environment utilise key computational knowledge services to assist the writing task, and to update the knowledge-bases when the writing task is completed (for example, new proposals becoming part of an "institutional memory").

4.1 Current Prototype

Our modelling and development efforts to date, currently in the third cycle of our iterative development approach, have produced a coherent WiCKOffice environment in which several knowledge services are available to proposal authors. A *knowledge fill-in* service and *knowledge recall* service are motivated by the need to provide timely and convenient access to knowledge collated from multiple diverse sources, which would otherwise have to be manually 'looked up' from multiple sources on the institutional intranet and the wider Web. A third service, *in-line guidelines*, also assists recall by exposing guidelines and constraints captured from the design specification (the EPSRC guidance notes) that are relevant to the part of the proposal document currently being worked on, presenting them to the user via the Microsoft Office Assistant interface (Fig. 3).

4.1.1 Filling in Forms

The *knowledge fill-in* service assists the author in filling in the Je-SRP1 form. For example, the author can specify the (partial) name of the Principal Investigator and instruct the service to retrieve appropriate (in context) instances from the knowledge-base to automatically fill in the remainder of the required information.

The majority of the information required to provide an assisted knowledge fill-in service for the Je-SRP1 form is already provided by the AKT Reference Ontology (our *research* ontology). However, leveraging this service is not

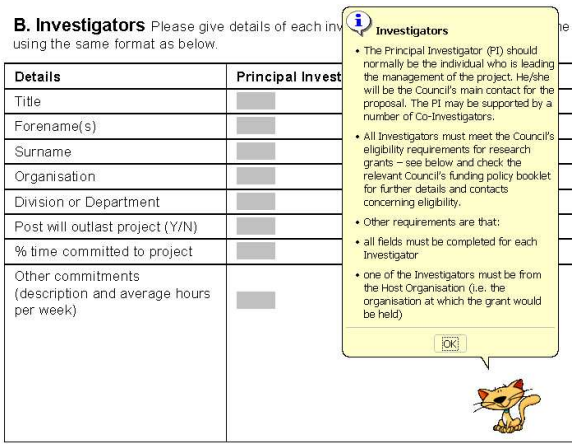


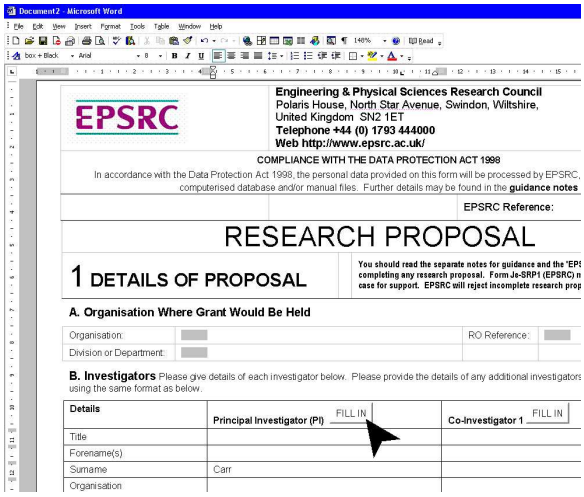
Figure 3: In-line guidelines presented via the Microsoft Office assistant.

as simple as filling each part of the form with an appropriate instance selected from the research ontology — different parts of the Je-SRP1 form "share" data about the same concept. For example, information relating to the Principal Investigator must entered in three different locations: section 1B (page 1) requires the PI's title, name, organisation, department, and commitments to other projects; section 2B (page 12) requires the PI's name (for the proposal declaration); and section 3B (page 13) requires the PI's contact telephone number, email address, fax number, *etc.*

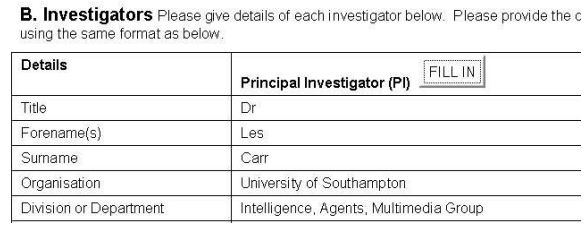
We have therefore used Microsoft Office 2003's new "smart documents" feature to add *semantic structure* to the otherwise unstructured Je-SRP1 template in the form of an XML Schema derived directly from the document ontology. The XML Schema identifies each 'sub-form' of the Je-SRP1 and groups together related sub-forms (thus, for example, describing the fact that information about the PI is shared by sub-forms 1B, 2B, and 3B). Each individual form field is marked up with three attributes — the ID of the sub-form to which the field belongs, a boolean value indicating whether that field is a preferred search field (in the case of the Je-SRP1, the PI's first name and surname are good search terms for a person instance in the research ontology; knowing the PI's title may not so helpful), and finally a `filled-in-by` attribute which identifies the slot of the matching knowledge instance which should be used to actually provide a value for the field.

When the author partially fills in a sub-form (Fig. 4a) and presses the "Fill-In" button, the XML structure of the document is consulted to determine which fields are part of the current sub-form (and also which fields are part of other sub-forms that share data with the current sub-form). Fields in the current sub-form with an `is-search-field` attribute value of `true` are then used by the knowledge fill-in service to construct an RDQL query to extract matches from the research ontology. In the case that multiple instances match the query, these instances are presented to the author who chooses the appropriate match. Finally, the `filled-in-by` attribute is used to map the slot values of the returned instance (which of course may originally have been harvested from multiple sources) to each associated field (Fig. 4b).

We have already noted that the EPSRC, in conjunction with several other UK research councils, has recently rolled



a. Author fills in partial details.



b. All sub-forms sharing data with current sub-form are populated from matching instance.

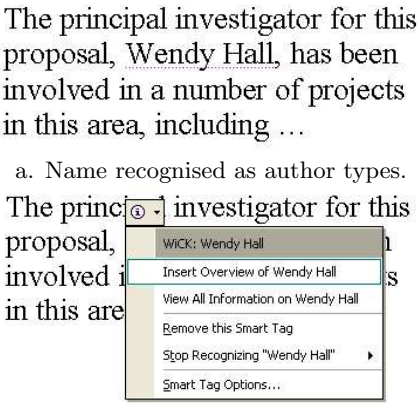
Figure 4: Using the knowledge fill-in service to help complete the Je-SRP1 form.

out its own assisted form filling system, the Je-S e-form[5], which provides some equivalent functionality to this service. Provided that each party has previously registered their details with the system, the author can select the host organisation, principal and co-investigators, referees and other staff from checklists and then download a partially completed Je-SRP1 form which contains all the required details of the selected parties, but still requires some unaided 'mandraulic' effort to complete in full. By contrast, we believe that the WiCKOffice approach of leveraging the functionality of multiple services operating over diverse *knowledge* sources (including, but not restricted to, employee data and information harvested from personal webpages and online directory services) not only allows authors to be aided in filling in *all* aspects of the Je-SRP1 form but also potentially offers wider applicability (adding new types of form requires only that form's semantic structure be elicited according the document ontology) than a *data*-based application.

4.1.2 *Knowledge in the Right Place at the Right Time*

The *knowledge recall* service assists the author in quickly and conveniently recalling appropriate knowledge from the research environment. Example (contextual) queries include "what papers relevant to this proposal have been published recently?", or "what relevant projects has this person worked on?". In response to such queries, appropriate knowledge from the knowledge-bases is selected and inserted directly into the document in the form of 'potted' summaries.

[5] `https://je-s.rcuk.ac.uk/`



a. Name recognised as author types.

b. Available actions in Previous Research section.

c. Available actions for recognised text "Wendy Hall" in References section.

Figure 5: Using the knowledge recall service, via the WiCKOffice Smart Tag.

As with the *knowledge fill-in* service, the AKT Reference Ontology provides the majority of knowledge utilised by this service. In the current implementation, given the name of a recognised person, project or place, the knowledge recall service assists the writer in recalling facts about it. We have seen that recent incarnations of Microsoft Office already provide a mechanism for recognising terms and presenting available "actions" associated with that term to the user in the form of Smart Tags. However, in the case of the Case for Support document, the author's information requirements depend on the section or part of the document currently being worked on. For example, the author might expect that typing "Les Carr" in the *Previous Research* section would make available options to "auto-summarise" or browse those facets of Les Carr's previous research history most relevant to the current proposal, whereas typing "Les Carr" in the *References* section would make available options to insert Les Carr's most recent and relevant publications, and typing "Les Carr" in the *Researcher Curriculum Vitae* section would make available options to insert a "mini CV" with information appropriate to the proposal (with links to knowledge sources in each case). However, prior to the release of Microsoft Office 2003, the actions made available through Smart Tags have been static; Office 2003 allows the set of available actions to be determined dynamically when the author activates (clicks on) a Smart Tag [13].

An XML Schema derived from the document ontology is again used to make explicit the structural semantics of the Case for Support document. When the author activates a WiCK Smart Tag by clicking on a highlighted term in the text, the XML structure of the document is consulted to work out which part of the document the text appears in (*e.g.* Background, References) and the actions offered by available services which are appropriate to the type of knowledge required in that section are presented (Fig. 5). We therefore describe this service as providing knowledge in the right *place* (*i.e.* appropriate to the author's current location in the document) at the right *time* (when a name of a recognised person, place or project is typed by the author).

4.2 Planned Future Services

Two further knowledge-based services are currently under development within the project proposal writing scenario. Using an appropriate proposal ontology, an *augmented experience* service provides the author with access to the "institutional memory" of previous research proposals, thereby augmenting the author's own experience of proposal writing ("what works? what doesn't work?"). For example, the author is assisted in evaluating the most important beneficiaries of the proposed research by being shown the beneficiaries put forward by other proposals (with an indication as to whether those proposals were subsequently approved or otherwise).

An *assisted writing* service attempts to assist the author in making higher-level decisions about relevant content to include in the proposal by suggesting appropriate instances from the subject ontology (for example, relevant projects, papers, resources) based on both the writing context and the text that the author has already written. For example, this service could use an internal bibliometric reasoning engine to detect that although the author has referred to a number of *knowledge acquisition*-related projects in the *Background* section[6] of the Case for Support, one statistically significant project has not yet been mentioned, and so offer to create a summary of the project from the relevant instances in the knowledge-base (gathering details of key personnel and publications) and insert the information into the appropriate sections of the Case document. Alternatively, by reasoning over a scholarly 'claim space' such as that facilitated by ScholOnto [16], the service could help the author formulate a convincing argument by providing access to, and ultimately inserting a summary of, selected claims in favour of and opposing the author's position.

4.3 Evaluation Against Requirements

Although the current WiCKOffice prototype represents only the first steps along the road to supporting a very complex process, our work to date already compares favourably to the requirements that we extracted from the results of our investigation of current practice. WiCKOffice's primary knowledge base, the AKT Triplestore, harvests its knowledge from a diverse range of information sources; WiCKOffice subsequently supports the effective management of these sources through the integration of toolbar and Smart Tag extensions with the popular writing tool Microsoft Word. To date these integrations include the toolbar-activated *inline-guidelines* and *knowledge fill-in* services and the Smart Tag-

[6]Guidance notes: "Demonstrate a knowledge and understanding of past and current work in the subject area both in the UK and abroad."

based *knowledge recall* service which allow authors to pull appropriate information from the knowledge base into the document.

By combining explicit descriptions the semantic structure of the proposal documents with dynamic Smart Tags, WiCKOffice is also able to support the different information requirements that arise from different tasks, even if those tasks are carried out within the framework of a single document, as with the Case for Support where the cutting-edge, historical context and work-plan tasks produce different sections of the document. Whenever a service inserts information into the document (for example, the *knowledge recall* service inserting a list of recent publications from a named author), appropriate knowledge markup is also inserted behind the scenes [5]. This markup, in conjunction with the document's explicit semantic structure, means that when the proposal is ultimately completed and submitted to the EPSRC for consideration, the knowledge can also be extracted and asserted into the triplestore for subsequent use by other authors.

5. CONCLUSIONS AND FUTURE WORK

Semantic Web activities are beginning to build large, flexible knowledge stores which can be leveraged for diverse purposes within an organisation. This paper has reported the latest efforts of a project which aims to assist authors in creating and re-using knowledge-rich documents within such an environment, specifically the preparation of a funding proposal. We have carried out an investigation to assess current practice within our department, and used the findings to inform a list of requirements for computer-based support for this task. The most arresting feature from this investigation was the diverse range of information sources that participants interacted with during the proposal writing process; there is a clear requirement therefore to support a large, diverse range of information sources. Furthermore, computer-based support for the proposal writing task should also assist the author in managing the complex interactions between these sources, increasing the author's effectiveness by minimising 'forced divided attention'. The investigation also revealed that there is a requirement to support different information sources and usage strategies in different tasks: each of the four different writing tasks that we identified within the proposal writing process showed a different pattern of information access. Finally, we observed a number of opportunities for information reuse in the proposal writing task, the most obvious being the range of strategies reported by participants in extracting information from previous project proposals. A final requirement therefore is that computer-based systems facilitate this evolutionary reuse cycle.

We have taken the first steps in integrating an office environment with knowledge-aware services to demonstrate how these requirements could be successfully met. Although we have yet to carry out a user evaluation of this proof of concept, as a case study in the application of Semantic Web technology to a specific business process it exercises the technologies very well, requiring:

- A widely applicable knowledge model.
- A fully populated, well maintained and evolving knowledge base.
- Appropriate and effective high level knowledge services, inferences and measures of trust.
- A familiar user interface to facilitate users in assisted knowledge creation.

The knowledge base we have leveraged within our writing environment is large (if incomplete), but without it, the requirements that we have established for building a computer-based system for supporting the proposal writing process from the ground up would be far too resource-intensive to justify; this work therefore gains much benefit from an existing knowledge store (in this case, the AKT Triplestore) and its established knowledge engineering processes (information harvesting *etc.*).

Our future work plans, aside from continued implementation of our integrated office environment, include a more detailed focus on the processes and mechanisms by which the knowledge provided by the AKT and WiCK knowledge-bases can be updated and maintained as more and more research proposals are produced. We also plan to carry out a systematic user evaluation using the participants of our initial fact-gathering investigation. Lastly, we are also working on a writing methodology for creating more complex, knowledge-rich documents such as multi-faceted Web sites and hypertexts.

6. ACKNOWLEDGEMENTS

This work has been funded in part by the EPSRC Knowledge Writing in Context (KWiC) project (GR/R91021/01) — now known as *Writing in the Context of Knowledge (WiCK)* — and the EPSRC Advanced Knowledge Technologies IRC (GR/N15764/01) in the UK.

7. REFERENCES

[1] The AKT Reference Ontology. `http://www.aktors.org/publications/ontology`, 2002.

[2] R. V. Benjamins, D. Fensel, and A. G. Perez. Knowledge management through ontologies. In *Proceedings of the Second International Conference on Practical Aspects of Knowledge Management, Basel, Switzerland*, 1998.

[3] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May 2001.

[4] L. Carr, S. Bechhofer, C. Goble, and W. Hall. Conceptual Linking: Ontology-based Open Hypermedia. In *Proceedings of the Tenth International World Wide Web Conference, Hong Kong*, 2001.

[5] L. Carr, T. Miles-Board, A. Woukeu, G. Wills, and W. Hall. The case for explicit knowledge in documents. In *Proceedings of the 2004 ACM Symposium on Document Engineering (DocEng2004), Milwaukee, USA*, pages 90–98, 2004.

[6] M. Dzbor, J. B. Domingue, and E. Motta. Magpie: towards a Semantic Web browser. In *Proceedings of the 2nd International Semantic Web Conference, Florida, USA*, pages 690–705, 2003.

[7] Engineering and Physical Sciences Research Council (EPSRC). Current grant portfolio. `http://gow.epsrc.ac.uk/ListOrganisations.ASPx?Mode=Inst&Order=VD`, Nov. 2004.

[8] C. Fillies, G. Wood-Albrecht, and F. Weichardt. A Pragmatic Application of the Semantic Web using SemTalk. In *Proceedings of the Eleventh International World Wide Web Conference, Honolulu, Hawaii, USA*, pages 686–692, 2002.

[9] S. Handschuh and S. Staab. Authoring and Annotation of Web Pages in CREAM. In *Proceedings of the Eleventh International World Wide Web Conference, Honolulu, Hawaii, USA*, 2002.

[10] S. Harris and N. Gibbins. 3store: Efficient Bulk RDF Storage. In *Proceedings of the 1st International Workshop on Practical and Scalable Semantic Systems (PSSS'03), Sanibel Island, Florida*, pages 1–15, 2003.

[11] J. Heflin, J. Hendler, , and S. Luke. Reading between the lines: Using shoe to discover implicit knowledge from the web. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, WI*, 1998.

[12] G. Hughes and L. Carr. Microsoft Smart Tags: Support, ignore or condemn them? In *Proceedings of the ACM Hypertext 2002 Conference, Maryland, USA*, pages 80–81, 2002.

[13] C. Kunicki. What's New with Smart Tags in Office 2003. *MSDN OfficeTalk*, 2003. Available from `http://msdn.microsoft.com/library/en-us/dnofftalk/html/office01022003.asp`.

[14] S. Lawrence, C. L. Giles, and K. Bollacker. Digital libraries and autonomous citation indexing. *IEEE Computer*, 32(6):67–71, 1999.

[15] T. Leonard and H. Glaser. Large scale acquisition and maintenance from the web without source access. In *Proceedings of Workshop 4, Knowledge Markup and Semantic Annotation, K-CAP 2001*, pages 97–101, 2001.

[16] G. Li, V. Uren, E. Motta, S. Buckingham-Shum, and J. Domingue. ClaiMaker: Weaving a Semantic Web of research papers. In *Proceedings of the 1st International Semantic Web Conference, Sardinia, Italy*, June 2002.

[17] H. Lieberman and H. Liu. Adaptive Linking between Text and Photos Using Common Sense Reasoning. In *Proceedings of the Conference on Adaptive Hypermedia and Adaptive Web Systems, Malaga, Spain*, pages 2–11, 2002.

[18] T. Miles-Board. *Everything Integrated: A Framework for Associative Writing in the Web*. PhD thesis, School of Electronics and Computer Science, University of Southampton, Southampton, UK, 2004.

[19] T. Miles-Board, C. Bailey, L. Carr, and W. Hall. Building a Companion Website in the Semantic Web. In *Proceedings of the Thirteenth International World Wide Web Conference, New York, USA*, pages 365–373, 2004.

[20] ontoprise GmbH. OntoOffice Tutorial. `http://www.ontoprise.de/documents/tutorial_ontooffice.pdf`, 2003.

[21] m. c. schraefel, Y. Zhu, D. Modjeska, D. Wigdor, and S. Zhao. Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections. In *Proceedings of the Eleventh International World Wide Web Conference, Honolulu, Hawaii, USA*, May 2002.

[22] N. R. Shadbolt, N. Gibbins, H. Glaser, S. Harris, and m. c. schraefel. CS AKTive Space or how we stopped worrying and learned to love the Semantic Web. *IEEE Intelligent Systems*, 19(3):41–47, 2004.

[23] M. Tallis. Semantic Word Processing for Content Authors. In *Proceedings of the Knowledge Markup & Semantic Annotation Workshop, Florida, USA*, 2003. Part of the Second International Conference on Knowledge Capture, K-CAP 2003.

[24] M. Tallis, N. M. Goldman, and R. M. Balzer. The Briefing Associate: Easing Authors into the Semantic Web. *IEEE Intelligent Systems*, 17(1), 2002.