ReSIST: Resilience for Survivability in IST

A European Network of Excellence

Contract Number: 026764

Deliverable D10: Prototype knowledge base

Report Preparation Date: December 2006
Classification: Public

Contract Start Date: 1st January 2006
Contract Duration: 36 months
Project Co-ordinator: LAAS-CNRS
Partners: Budapest University of Technology and Economics
City University, London
Technische Universität Darmstadt
Deep Blue Srl
Institut Eurécom
France Telecom Recherche et Développement
IBM Research GmbH
Université de Rennes 1 – IRISA
Université de Toulouse III – IRIT
Vytautas Magnus University, Kaunas
Fundação da Faculdade de Ciencas da Universidade de Lisboa
University of Newcastle upon Tyne
Università di Pisa
QinetiQ Limited
Università degli studi di Roma "La Sapienza"
Universität Ulm
University of Southampton

Contents

1 - Summary

The Description of Work states:

> **Deliverable D10: Prototype Knowledge Base, developed by task IT-T1**
>
> This deliverable will take the form of a demonstration system providing a prototype on-line information service on resilience building technologies, embodying initial versions of dependability and security ontologies, and acting as a repository of information on relevant people, projects and publications.

It further states:

> **Task IT-T1: Resilience Knowledge Base (RKB)**
>
> This task aims to provide convenient integrated access to extensive information built up from across Europe -- specifically about people, projects, institutions, events, publications, etc., of relevance to system resilience researchers in both the dependability and the security domains, and to their sponsoring organisations. The work will be structured into four sub-tasks addressing Acquisition, Ontology Engineering, Information Storage, and Information Publishing, respectively. The work will primarily be undertaken by Southampton, but all partners will contribute input and validation activity. Furthermore, to ensure that each partner's work and interests are well represented in the RKB, every partner will have a representative on an associated Editorial Board providing oversight and guidance for the development and content of the knowledge base.
>
> Each of the four sub-tasks will first conduct a 12-month preliminary phase which will include validation of the prototype knowledge base as fit for purpose through the use of test scenarios developed by the partners. The validated prototype knowledge base at this point will form the first deliverable, D10, from WP1. A second, six-month, phase will then be carried out to consolidate, enhance and extend the work from the initial phase.
>
> *Acquisition*: The preliminary phase will accumulate information encompassing all of the network partners, initially concentrating on people, projects and publications in the network but also gathering data on related activity, as well as drawing on existing information resources at Southampton. Steps will also be taken to initiate data gathering from a much broader set of sources with relevance to resilience; examples of such sources are : RISKS, CORDIS, funding agencies, relevant intra- or extra-European projects (e.g., related US projects). In the second phase, the availability of an initial ontology will assist the harvesting of more and better data from the partners, and in extending our reach to wider sources.
>
> *Ontology engineering*: Definition of a resilience-specific ontology will be initiated during the preliminary phase, building on existing taxonomies and standards. Southampton, Newcastle, LAAS and Kaunas will be directly involved, but the development of the ontology will draw on discussion and review from all partners. In the second phase, further refinement of the ontology will be informed by its role in supporting data acquisition, and by engagement with task IT-T2 on formal descriptions of meta-data.
>
> *Information storage*: Primary infrastructure for the RKB will be established using a Project Triplestore (an open source knowledge base technology) which we estimate will hold at least 106 triples (basic information entities) by the end of the preliminary phase. Bibliographic data held in the RKB will be mirrored from an ePrints archive, thereby gaining considerable leverage from a rapidly expanding digital archiving phenomenon. At M18 we anticipate that the Triplestore will have accumulated c. 2x10^7 triples and there will be a need for ongoing maintenance and support.
>
> *Publishing*: During the preliminary phase a demonstration RKB, with similarities to CS AKTive Space, will be made available to the partners for validation in terms of facilities and content. This will provide a prototype on-line information service on resilience technology to the partners, and thereby advance our shared understanding of capability and scope. In particular, this will enable the selection of an effective set of visualisation capabilities that will be incorporated to augment the RKB during the second phase.

Thus, the primary component of deliverable D10 is the prototype RKB that has been developed, and is available at http://resist.ecs.soton.ac.uk/explorer/. This present document serves to characterise the prototype in terms of its organisation, content, and mode of access.

2 - The Deliverable

The main prototype user interface to the RKB (the Explorer interface) can be found at http://resist.ecs.soton.ac.uk/explorer/ (details of this interface, in the form of a User Guide, are given in Section 3, and a detailed report on the evaluation exercise that was undertaken during December 2006 is provided as Appendix 1)

A raw triple browser can be found at http://resist.ecs.soton.ac.uk/browse/ - for further details see Information Storage section below.

SPARQL and RDQL query interfaces can be found at http://resist.ecs.soton.ac.uk/sparql/; this exposes the contents of the RKB in a form that can be used by external applications.

The Project's internal Wiki (http://resist.ecs.soton.ac.uk/wiki/) also uses the information in the RKB, but is not publicly available. Additional experimental geographical-style user interfaces to the RKB, providing convenient access to information about ReSIST partners and courseware (See Figures 1.1 and 1.2), have also been provided.

It should be noted that the start of the RKB development activity was compromised by the requirement to begin the contract at exceptionally short notice. Consequently no preparation could be done, and it was not possible to appoint the staff in good time. Nevertheless, exceptionally good progress was made in developing and populating the RKB, in designing and implementing its interfaces, and in performing a first evaluation of the RKB Explorer.

We here comment further on the following parts of the Task.

Acquisition

General information sources were evaluated for the value they brought versus the cost of the effort to acquire and maintain them. We were in fact able to acquire all the data held in CORDIS (on EU research projects), by NSF (US research projects), in DBLP, and Citeseer (ResearchIndex) and by the ACM, on publications, and on Dependability problems from the RISKS Index of "Computer-Related Risks to the Public". In total over 40 million information items ('triples' or 'facts') have thus already been collected and integrated into the knowledge base.

It might seem that gathering data, as we have, from a wide range of resources is less useful than concentrating on resources that we already know are resilience-oriented; this is not the case. An expected value of a system such as the RKB is in giving the user relevant material from unexpected sources. In particular in the case of resilient systems, for example it is good to include resources from application domain sources, where the emphasis of the source relates to the application, and might not be seen as primarily a resilience issue.

We were able to gather data (on people, publications and projects) from almost all the partners, although the task proved more challenging than expected, since not all of the partners make such information available, for example via their Web sites, or in convenient database formats. However, the relevance of this source was reduced, as the other sources that we obtained were more successfully harvested than expected.

It was decided to use a Wiki to facilitate project interaction, but it was considered highly desirable that public information in the Wiki should be reflected in the RKB and *vice versa*. A Semantic Wiki was therefore used, and adapted to interface with the RKB. In particular, a specialist page was built in the Wiki to facilitate the provision by partners of details of their research interests (see Figure 4.3).

An ontology was defined to support representation of information about Courseware and a further specialist page was provided in the Wiki (Figure 1.3) to facilitate the acquisition of this information from partners directly into the RKB. Furthermore, this interface permits project members to browse available courses, their aims, prerequisites and requirements, and to easily maintain these details.

The Description of Work suggested we would acquire around one million triples by the end of this preliminary phase. RDF 'Triples' are the basic unit of information storage used to provide an integrated knowledge base - an example of a triple which represents a single fact being <ReSIST, has-member, Jean-Claude Laprie>. In practice we have found that the resources above amounted to approximately 40 million triples.

Ontology Engineering

The work done in this area is described in detail in Section 4 of this document.

Information Storage

The Project Triplestore has proved highly satisfactory for the requirements of storing the large volumes of RDF data gathered. The provision of the SPARQL endpoint permits queries to be submitted and information to be retrieved from the repository in an efficient manner, and it is through these means that the various prototype interfaces have been implemented.

The ability to harvest from a wide range of resources (CORDIS, CiteSeer, etc.) meant that the problem of co-reference resolution was more significant than would otherwise have been the case. Co-reference refers to the situation in which data sources use simple text to refer to resources, or their own URI. The RKB then has the non-trivial problem of detecting duplicate variant references (e.g. "Jean-Claude Laprie" and "J.-C. Laprie"), and deciding whether such text strings and URIs actually refer to the same entity in the real world. The problem is particularly acute for publications and people. We therefore researched, developed and have implemented an automatic system for identifying such co-references, storing the information and exposing it through a web service. This is then used within the prototype interface to combine data from disparate sources, and to remove unwarranted duplicate references to individual resources. This system will now allow us to add further data sources without exacerbating the problem.

Publishing

The proposal for a CS AKTiveSpace-style interface as a means of 'publishing', i.e. accessing the contents of the RKB, was reviewed. While such a geographical interface was considered satisfactory, it was considered that the main visualisation would be better if it gave a view of the relationships between entities in the RKB, something that was done by dynamically calculating and displaying what are termed "Communities of Practice". Geographical visualisation, although exciting to see, is in general less useful than subject relationships, especially as the huge amount of data would be hard to show effectively. However, geographical visualisation (based on Google Maps) was nevertheless provided for specific needs: ReSIST partners (Figure 1.1) and the places courses are taught (Figure 1.2).

The main RKB Explorer interface is described in detail in Section 3, the User Guide.

Review

The RKB Explorer and the RKB content were reviewed by the members of the RKB Editorial Board and others during December. The results of this review can be found in Appendix 1. The prototype system was in general very well-received and performed reasonably well. Many comments have already been acted on where possible, and

others noted for future enhancements although it will not be possible to address all issues with the resources currently planned.

Figures

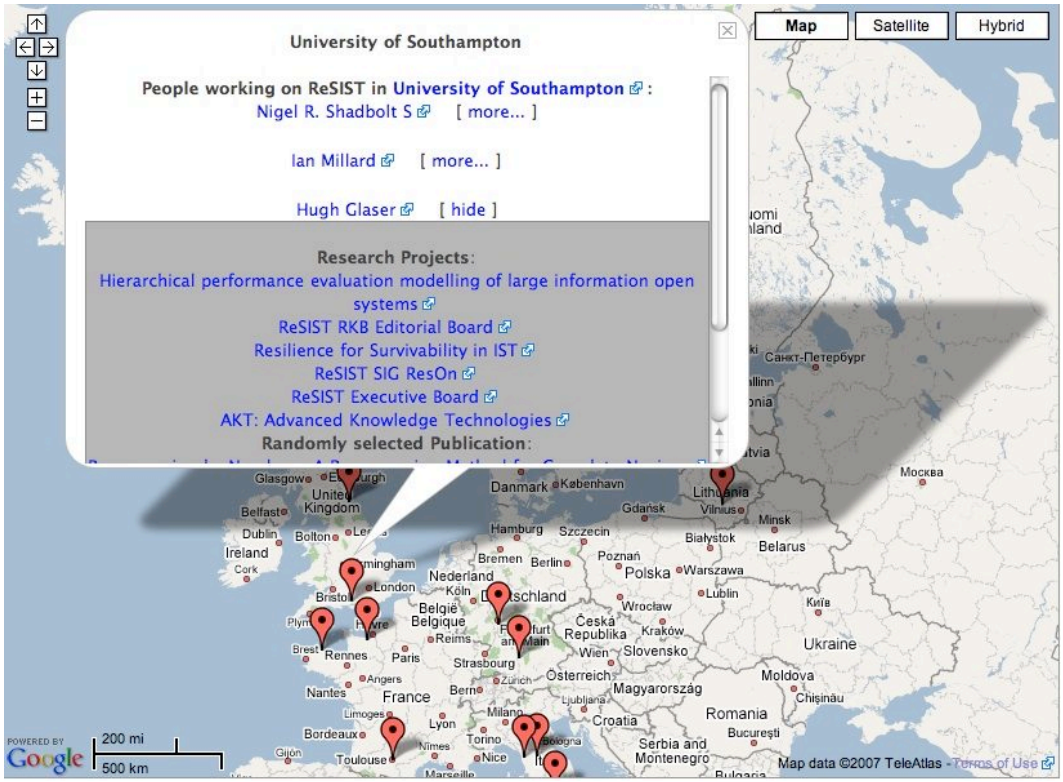


Figure 1.1 – Project partner locations and details screenshot
http://resist.ecs.soton.ac.uk/wiki/maps/resist-members

Figure 1.2 – Course locations and details screenshot
http://resist.ecs.soton.ac.uk/wiki/maps/courseware

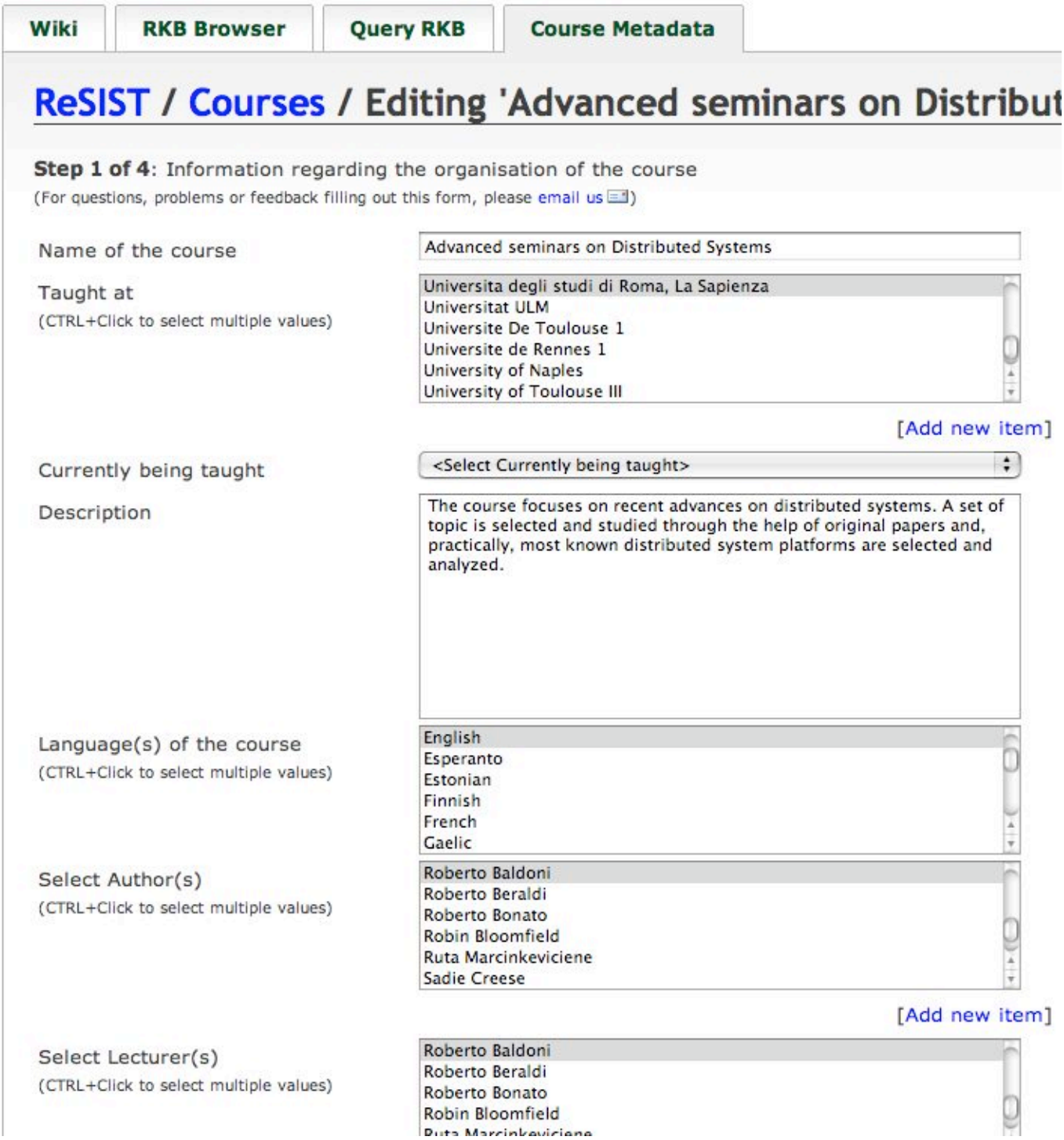


Figure 1.3 – Entry Form for Courseware Information
http://resist.ecs.soton.ac.uk/courseware/

Walkthrough of RKB Explorer functionality

The RKB contains a significant quantity of raw information, gathered from disparate sources and encompassing a wide range of Dependable Systems and more general computing topics. While SPARQL endpoint and Triple Browser interfaces permit direct access to the facts held within the repository, the RKB Explorer interface has been designed to enable useful interaction by every-day users who do not have detailed Semantic Web expertise.

Information held within the RKB is refined by the Explorer interface before it is presented to users, through the application of advanced co-reference resolution algorithms and community of practise analysis to consolidate duplicate references and to identify related resources. Users may search and browse through the information available based around the four core themes which are present within the Explorer interface, namely *People*, *Research Topics*, *Publications* and *Projects*. At any one time the top half of the interface window details an instance of one of these types of resource, while the lower half lists those resources of each type which are related to the currently selected item.

Upon first visiting the RKB Explorer, a user finds the interface focussed on the ReSIST project within the graph display. This is indicated by the title at the top of the graph, and by the fact that the central ReSIST node is highlighted in red. Related resources of the same type as that highlighted (projects) are shown as linked nodes within the graph, while details of the selected project are presented within the panel to the top right. The four panels in the lower half show those people, research areas, publications and projects which are related to ReSIST.

Selecting a related item, for example the person named “Alberto Pasquini”, will change the display to focus on that resource. Again, the focussed item is detailed in the upper half, while the lower half shows related resources. Double-clicking on another person such as “Roberto Bonato” will expand the relevant node within the graph display, with the added benefit of identifying those people who are common to both Alberto and Roberto.

A user may continue browsing the current group of people, viewing each person’s publications, projects and research areas (where available), or change the entire display by selecting an alternative type of resource to view, such as a paper or project. Such ‘progressive’ browsing enables a user to discover further information related to what they are viewing, with associations determined by detailed analysis over the underlying knowledge base.

At the top of the screen is a menu bar, indicating which of the four types of resource is currently being viewed. Alternatively, from here a user may select the Search option, enabling a textual search to be performed on the contents of the knowledge base. Providing a search interface brings the important facility to enable users to begin browsing from a specific starting point, be it searching for a person’s name, or a keyword which may return related publications, research areas, or projects. Results matching the search term are divided into the four typed lists, and selecting an individual result returns the user to the familiar graph display and related resources.

The following section provides more detailed descriptions of the various controls and functionality of the RKB Explorer.

3 - RKB Explorer User Guide

The text in this section is taken from the RKB Explorer's "Help" page

Overview

The RKB Explorer is started by going to the URL: http://resist.ecs.soton.ac.uk/explorer/.

After loading the page and Java Applet, you may be asked if you trust the security certificate. A negative answer is perfectly acceptable: it will however mean that you will be unable to copy the graph pane to your clipboard.

It is intended that the certificate will be properly signed in due course.

The interface is now ready for use.

The user can focus on an item in one of four dimensions:

People	Individual people in the RKB. They may have multiple identities (B. Randell and Brian Randell), but the closer they are to Resilient Systems the better their identities will be curated.
Research Areas	In principle, any area from Computer Science; however, those topics related to Resilient Systems are better curated.
Publications	Data from a wide range of Computer Science resources, including all of CiteSeeer and DBLP.
Projects	Projects that the RKB has discovered. In particular, the RKB is up to date on all projects funded by the European Union (EU) via the CORDIS database.

At any given time, the upper half of the display details information about a single resource - the current selection - which may be of any of the four types. To the upper right is a panel providing detailed information regarding the current resource, while the largest part of the interface to the upper left is a graph representation of the current resource, shown in the context of related resources of the same type (a so-called "Community of Practice" graph). The graph permits the user to change the current focus, and expands on double-click to show additional resources as required. These relationships are 'calculated' dynamically from the heterogeneous large sets of information about people, publications, research topics and projects that have been integrated into the RKB.

The lower half of the RKB Explorer displays four columns detailing the people, research topics, publications and projects which are related to the currently selected resource. Selecting any of these items changes the focus of the entire display to reflect the new resource. If the newly selected resource is of the same type as the current resource, then it is added to the graph display, otherwise the graph is cleared and replaced with the new representation.

For example, if the item of focus is a Project, the graph view will show projects in similar areas, the People list will show people associated with the project, and Publications will show project-related and field-related publications.

If the system appears to be taking a long time (> 30 seconds) without populating the lists, the interface can be returned to its original state by pressing your browser's reload button.

Navigation

In normal use, focus is changed by choosing an object from one of the dimensions, either in the graph view, or in the list view. In addition, a major focus change can be made by clicking on one of the menus at the top.

Two actions can be taken: a single click simply changes the focus. A double click causes the RKB Explorer to add related items to the graphical view.

In the menu area there is a “Search” choice, which allows the user to search for words or names in the whole of the RKB. When used, the interface switches to show results in all these items.

There is also a “Recently Viewed” choice, which takes the user to a previously-selected item in the RKB.

With the exception of the controls for the Graphical view, the interface remains static as the focus changes.

User Controls

The interface window provides control components of four types:

Menu Bar

The Menu Bar has the following components:

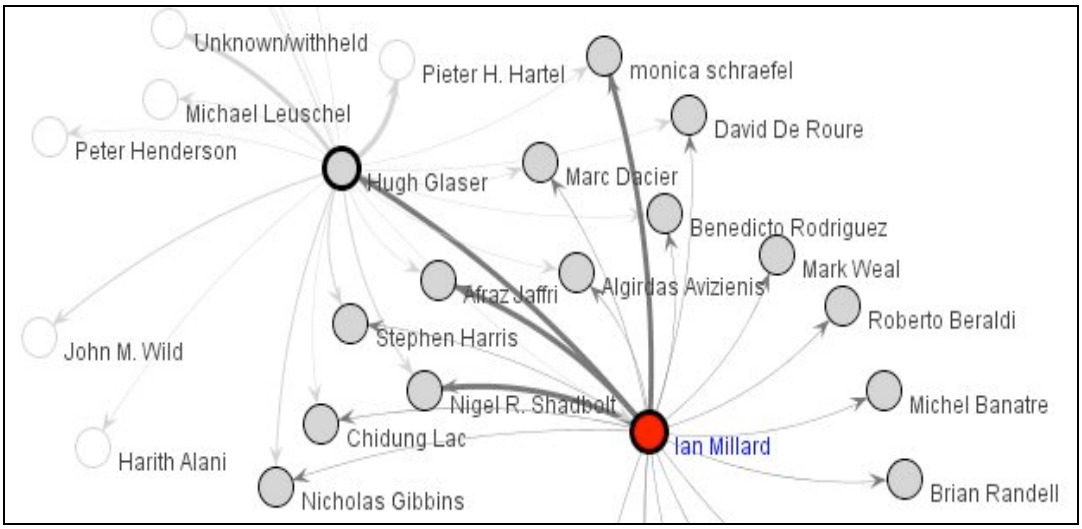
People	Make the view of the Graph focus on people
Research Areas	Make the view of the Graph focus on research areas
Publications	Make the view of the Graph focus on publications
Projects	Make the view of the Graph focus on projects
Search	Brings up a text search box. The entered text is then used to search all fields, with the results being sorted by their type into the four panels below. Names within the RKB are in the form “given-name family-name”. Note that the exact phrase entered is searched within the RKB. Regular expression matching can be used.
Recently Viewed	A list of recently viewed items can be seen. Selecting an item makes the item the focus and adds it to the existing graph.
Help	Brings up the help window.

Using the top menu to select one of the four types of view will change the display to show either the last person, research area, publication, or project that was viewed by the user, or a suitable ReSIST related default if no resources of the selected type have yet been shown.

Similarly, selecting Search will present the last search results, if available.

The item on the menu bar with the highlighted thick border indicates the currently selected type of the display.

Graph



The Graph display gives a visual presentation of the item of focus, and indicates those items of the same type to which the current item is related. Thus in the figure, "Ian Millard" is in focus, and the node is highlighted in red. The directly related people are attached by links, where the width of the link (weight of the line) indicates the strength of the connection. Directly connected people are shown in grey, and people who are not directly connected are left white. Those for whom the related items are shown are depicted by a dark black edge to the circle. (These relationships have been dynamically discovered, or 'calculated', taking into such factors as co-membership of projects, and co-authorship of papers.)

The top left of the pane shows the name of the item in focus, and the top right has a link which makes the graphic pane use more of the screen. There are scroll bars to move around the graph.

There are a variety of ways of manipulating the view given by the graphic pane.

Clicking on a node has the effect of changing the focus to that item; double clicking changes the focus, but also brings up the related items in the graph window.

The buttons on the right function as follows (there are explanatory 'tooltips', but at present they only function when the graph has the focus):


Zooms in on the graph


Zooms out from the graph

Zooms to (1:1), a natural size


Zooms to make the graph fit in the window


Brings up the Help Page

Brings up a scroll list from which any member of the ReSIST Project can be selected for the new focus. Only available when the graphic view is being used to view people.

Remove selected nodes

17 / 16 The number of nodes and number of links in the graph.

Copy the graph pane to the clipboard - only permitted if the security certificate has been trusted.

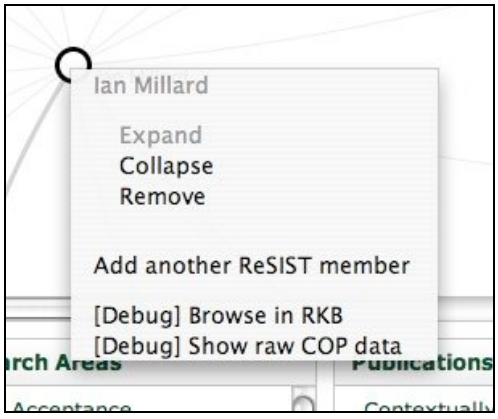
Finally, context menus and key modifiers are available:

Control-click

Control-clicking on a node causes that node to become the centre of the pane, otherwise does nothing.

Right-click or Alt-click (Mac)

Right- or Alt-clicking on a node brings up a context menu:



Using the context menu on a node allows the user to Expand, Collapse or Remove the node and related items as appropriate. Can also bring up the scroll list from which any member of the ReSIST Project can be selected for the new focus.

Various debug options are available from this menu, but not intended for general use.

Right-click or Alt-click (Mac)

Right- or Alt-clicking on the graph anywhere other than on a node brings up a different context menu:

Using the context menu when the mouse is off a node allows the user to bring up the scroll list from which any member of the ReSIST project can be selected for the new focus.

Various debug options may be available from this menu.

Detail pane

The detail pane gives appropriate information about the item that is the current focus.

Related lists

This part of the window seeks to provide information in the four categories that are related to the item of focus, according to connections in the RKB. The ordering of the items in each list aims to be in decreasing order of level and strength of connection.

To identify related resources, a number of complex analyses are performed over the data held within the RKB. Queries are executed to discover other resources which are related to the currently selected resource, based on pre-defined relationships such as co-authorship, or affiliations to a common project or organisation. These relationships are weighted, with each result contributing to an overall measure of ‘closeness’ for potential candidate resources.

The items in these lists can be clicked or double clicked. Clicking has the effect of changing the focus to that item; double clicking changes the focus, but also brings up the related items in the graph window.

System Requirements

The RKB Explorer is built to run on most modern operating systems and browsers.

It requires a standard installation of Java, version 1.4 or above.

You can check whether you have Java installed by going to http://www.java.com/en/download/help/testvm.xml

If it is not installed, follow the instructions to install it.

In particular, it has been used with the following software

Windows: Internet Explorer (v6, v7) and Firefox (v1.5, v2)

MacOS X: Safari (v2), Firefox (v1.5) and Camino (v1.0.3)

Linux: Firefox (v1.5)

Internet Explorer on **MacOS X** has not been maintained by Microsoft for a long time, and so is no longer able to support applications such as this.

4 – Resilience Ontology

ReSIST's Resilience Knowledge Base (RKB) makes use of a number of pre-existing ontologies, e.g. for general concepts such as projects, publications, etc.. The principal ontology that has been developed specifically by ReSIST is that on Dependability and Security, which is closely based on material in the paper "Basic Concepts and Taxonomy of Dependable and Secure Computing" (IEEE TDSC, Jan 2004), by Avizienis, Laprie, Randell and Landwehr. This paper contains a number of taxonomies, e.g. for types of failure, fault and error, so these have been extracted, and combined into just two taxonomies, one on Dependability and Security, and one on Systems (see Figure 4.1). These have been represented as an ontology, expressed in OWL, which incorporates 166 terms related to Dependability and Security, and 23 to Systems (see Figure 4.2).

One direct use of this ontology in the RKB is for allowing individual researchers to specify their particular research interests. A hierarchical form has been provided, via the ReSIST internal wiki, to make it very easy for all ReSIST members to specify their interests (see Figure 4.3) - in due course a similar facility will be provided for general use. At this stage the main use made of the information so provided is in the Research Areas pane in the RKB Explorer's "People" tab. This form, and the Details pane of the RKB Explorer's "Research Areas" tab, make use of the definitions of dependability and security terms that have also been extracted from Avizienis et al.

At present the largest body of material in the RKB that contains detailed technical metadata is that obtained from the ACM, providing details of papers appearing in the ACM Publications Library. All such papers when originally published specified the relevant ACM Computing Classification System categories. This information is accessible as metadata and has been captured for the RKB. Unfortunately, although the ACM classification is in general very detailed, the subjects of Dependability and Security are not well treated in it, being neither detailed nor conveniently grouped and located - a listing of correspondences between the ACM categories and the Dependability and Security ontology has therefore been produced. (This is intended to aid integration of the ACM Publications Library, already installed in the RKB Triple Store, with information obtained from other sources.)

Work has started at Kaunas on the automated processing, in order to identify meaningful sets of keywords, of a corpus consisting of the text of the entire set of papers that have been published in the annual IEEE Dependable Systems and Networks conferences and their predecessor conferences since 1971. This corpus, which constitutes one of the most important sources of information on dependability research, unfortunately does not provide ACM categories, or anything equivalent, in the way of useful technical metadata. Our aim therefore is to provide means of linking DSN papers to ReSIST's Dependability and Security ontology.

The present ReSIST Dependability and Security ontology is, as indicated above, already being made use of in the RKB. However, it is also intended to serve as a starting point for one to be developed describing the detailed characteristics of resilience-related tools and components in support of work on Resilience-Explicit Computing.

Resources permitting, one further avenue we are considering for the future is that of providing mechanisms that allow/encourage RKB users to associate their own choice of descriptive tags with items they find in the RKB, in order to obtain the combined advantages of an ontology and a folksonomy in categorizing and aiding access to the RKB's already large knowledge base.

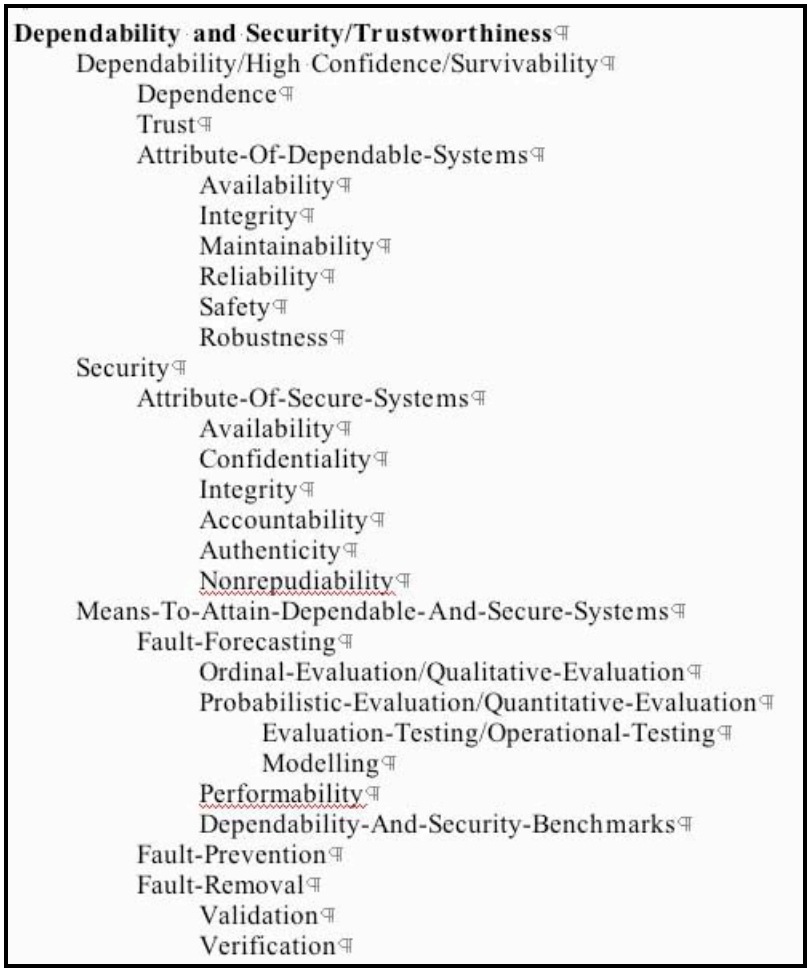


Figure 4.1 - The first section of the taxonomy derived from Avizienis et al.

```
- <rdf:RDF xml:base="http://resist.ecs.soton.ac.uk/ontologies/resist">
  - <owl:Ontology rdf:about="">
      <rdfs:label>ReSIST Ontology</rdfs:label>
      <dc:title xml:lang="en">ReSIST Ontology</dc:title>
    - <dc:description xml:lang="en">
        The ReSIST Ontology encompasses concepts within the fields of research in Resilient, Survivable and
        Dependable Systems
      </dc:description>
      <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string">0.2.5</owl:versionInfo>
      <dc:created>2006-11-28</dc:created>
      <dc:creator>ReSIST NoE</dc:creator>
    </owl:Ontology>
  - <rdf:Description rdf:about="http://resist.ecs.soton.ac.uk/ontologies/resist#Dependability-And-Security">
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
      <rdfs:label>Dependability And Security, Trustworthiness</rdfs:label>
    - <rdfs:comment>
        Two somehat overlapping concepts, with dependability being an integrating concept that encompasses the
        attributes: availability, reliability, safety:, integrity and maintainability, while security encompasses
        comfidentiality as well as integrity and availability.
      </rdfs:comment>
      <rdfs:subClassOf rdf:resource="http://www.aktors.org/ontology/portal#Research-Area"/>
    </rdf:Description>
  - <rdf:Description rdf:about="http://resist.ecs.soton.ac.uk/ontologies/resist#Dependability">
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
      <rdfs:label>Dependability, High Confidence, Survivability</rdfs:label>
    - <rdfs:comment>
        The original definition of dependability is: the ability to deliver service that can justifiably be trusted. The
        alternate definition, that provides the criterion for deciding if the service is dependable, is: the dependability
        of a system is the ability to avoid service failures that are more frequent and more severe than is acceptable.
      </rdfs:comment>
      <rdfs:subClassOf rdf:resource="#Dependability-And-Security"/>
    </rdf:Description>
  - <rdf:Description rdf:about="http://resist.ecs.soton.ac.uk/ontologies/resist#Dependence">
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
      <rdfs:label>Dependence</rdfs:label>
    - <rdfs:comment>
        The dependence of system A on system B represents the extent to which system A's dependability is (or
        would be) affected by that of System B.
      </rdfs:comment>
```

Figure 4.2 - The first section of the Dependability and Security Ontology, in OWL.

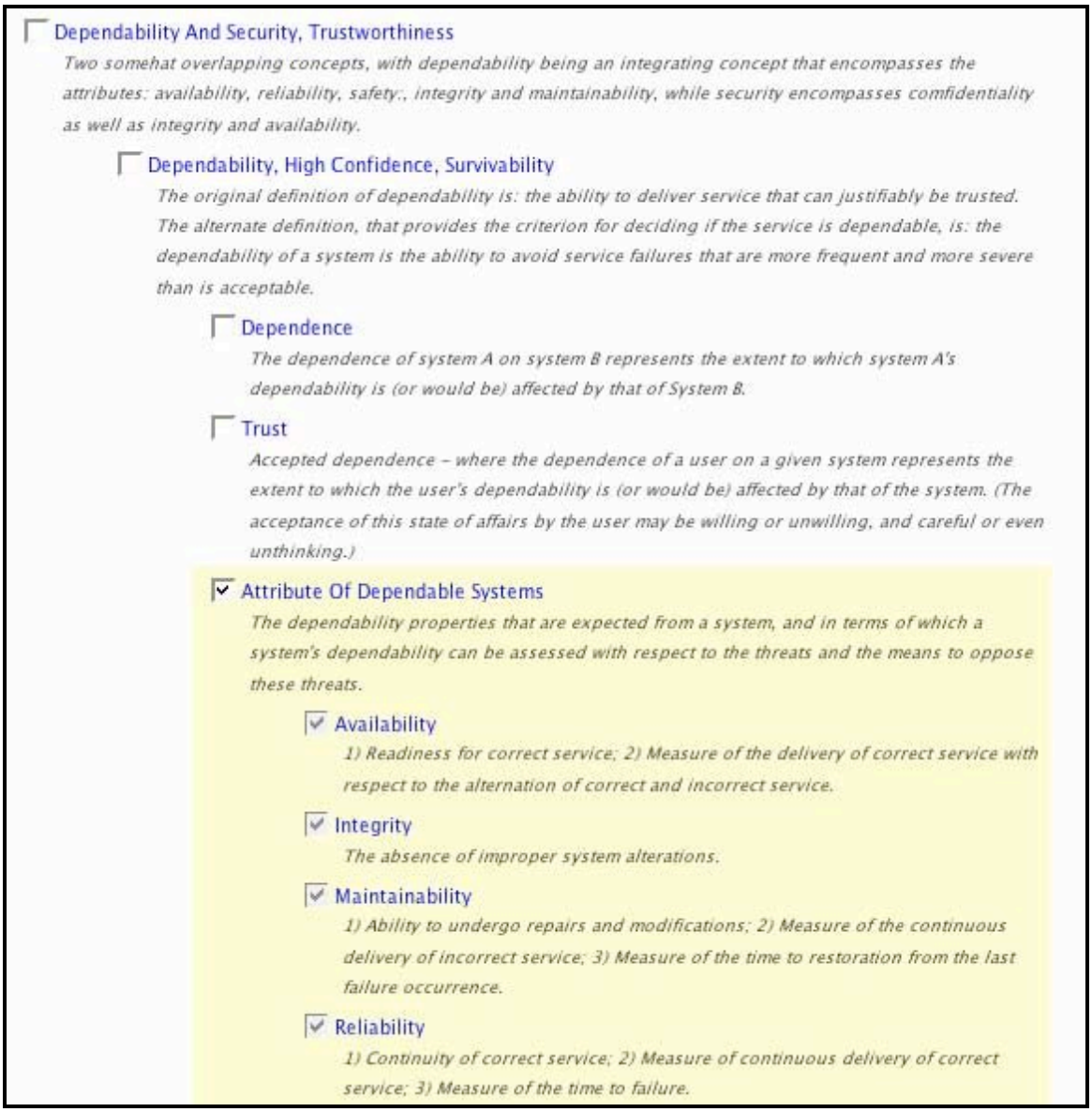


Figure 4.3 - The first part of the form provided for researchers to specify their interests.

Appendix 1 – Report on Evaluation of ReSIST RKB Explorer

Evaluation of ReSIST RKB Explorer

Ian Millard, Hugh Glaser

<icm, hg>@ecs.soton.ac.uk

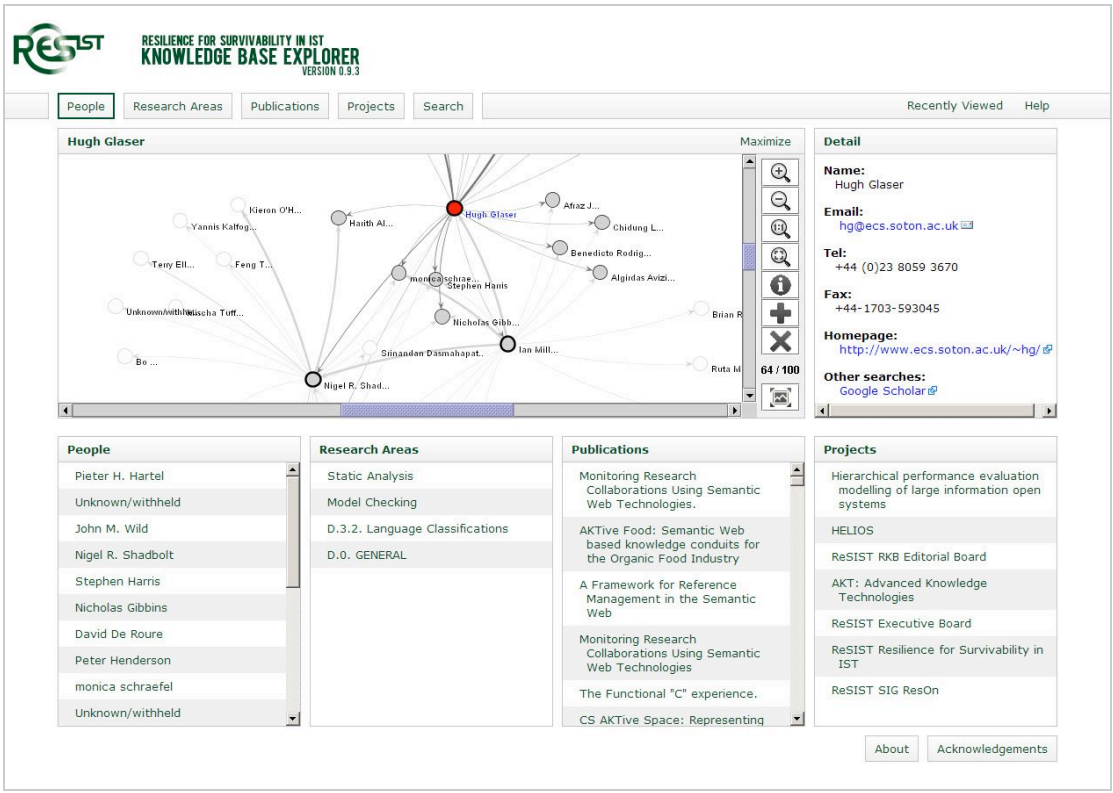
University of Southampton

Introduction

As part of the ReSIST Network of Excellence the development of an advanced semantic knowledge repository has been commissioned, capable of combining a wide range of information related to both the project and the more general field of resilient and dependable systems research. To enable users to access and interact with the information contained within the Resilience Knowledge Base, a prototype web-based interface has been developed. This document summarises the key experiences, issues and problems raised in a preliminary evaluation of this interface, which was carried out by a cross-section of the project membership.

RKB Explorer Overview

The RKB Explorer, shown below, enables users to browse information based around the four core themes of person, research topic, publication and project.



The upper half of the display details a single resource, the current selection, which may be of any of the four types. To the upper right is a panel providing detailed information regarding the current resource, while the largest part of the interface to the upper left is a graph representation of the current resource, shown in the context of related resources of the same type. The graph permits the user to change the current focus, and expands on double-click to show additional resources as required.

Meanwhile, the lower half of the RKB Explorer displays four columns detailing the people, research topics, publications and projects that are related to the currently selected resource. Selecting any of these items changes the focus of the entire display to reflect the new resource. If the newly selected resource is of the same type as the current resource, then it is added to the graph display, otherwise the graph is cleared and replaced with the new representation.

To identify related resources, a number of complex analyses are performed over the data held within the RKB. Queries are executed to discover other resources that are related to the currently selected resource, based on pre-defined relationships such as co-authorship, or affiliations to a common project or organisation. These relationships are weighted, with each result contributing to an overall measure of 'closeness' for potential candidate resources.

Summary of reviews

On the whole, the RKB Explorer was very well received by the project members who were asked to experiment with the interface and perform an initial evaluation; comments made were that it was intuitive, effective, and enjoyable to use. However, one must bear in mind that the RKB itself and the Explorer interface are only at a prototype stage, and require significant curation, maintenance and additional development before it should be considered an enterprise level facility.

Most of the comments and issues raised in the internal review fall into three categories –

- Shortcomings which are known by the development team and were expected to be raised. These issues have largely arisen as a result of limited time and resources available for the prototype development
- Unanticipated issues regarding the operation and/or user's understanding of the interface, which are to be addressed where possible
- Other comments pertaining to minor issues or individual preferences, which will mostly not be acted upon as they cannot be accommodated within the available resources.

The issues raised by the RKB Explorer reviewers are summarised in the following table, with accompanying descriptions below. A total of 24 individual reviews were received from ReSIST partners.

Issue		Reviews which raised issue																							
1	Multiple entries for same person/paper/project	✓	✓	✓	✓		✓			✓	✓	✓	✓			✓			✓	✓	✓	✓			✓	✓
2	Navigation / Go back / Undo Previous view	✓			✓		✓	✓			✓	✓	✓		✓				✓			✓			✓
3	Unclear how relationships are determined / linked in graph		✓					✓			✓		✓	✓			✓		✓						✓

4	Accented characters not displayed correctly	✓	✓			✓	✓	✓		✓														✓
5	Errors or inconsistencies in source data						✓		✓	✓	✓							✓			✓		✓	
6	Detail panes could show more information					✓	✓			✓		✓			✓			✓						
7	Ordering of names and publications seems random									✓		✓	✓							✓	✓			✓
8	Old browsers not supported / Layout problems				✓					✓			✓	✓										
9	Poor indication of activity when application is “busy”														✓	✓					✓			
10	Extraneous ontological entities (akt:Research-Area etc)						✓									✓								
11	Institutions poorly represented							✓													✓			
12	Top menu -- confusion of use/action			✓			✓																	
13	Missing associations with co-workers		✓																					
14	Distinguish between academics and students										✓													
15	Failed to find search															✓								
16	Firewall issues prevented java applet from loading																✓							
17	Improve results for scientific council members and reviewers									✓														
18	Live-ness of imported data										✓													
19	Names can appear differently in list, graph, and detail panel														✓									
20	ReSIST members and publications not distinguished							✓																
21	Search with hyphens																							✓
22	Selecting more than one person, union of two research areas									✓														
23	Search for Surname/Forename -vs- Forename/Surname			✓																				
24	Tool tips on graph controls																					✓		

25	Unclear of what resource/entity is the current focus	✓		
26	Unknown/withheld appearing as the name for certain people	✓		
27	Search results pagination		✓	
28	Search results confusion			✓

1. Multiple/duplicate entries for the same person/paper/project

Many users reported the problem of duplicate instances for the same person, project or publication being listed within the knowledge base. Through the utilisation of a common ontological model, the RKB has been able to combine a very large quantity of information, from a wide range of disparate information sources. As can be expected, in some cases there is considerable overlap between these resources, resulting in different descriptions of the same thing.

The task of co-reference resolution, or identifying and consolidating these multiple references, is arguably the most fundamental and most ignored problem within the semantic web vision. Only when the different fragments of information are joined up does the real power of bringing together the various sources become apparent, yet the methods and technologies to achieve this are still a subject of active research.

This challenge is of particular interest to the team at Southampton, who have devised a system to attempt to tackle the co-reference issues. Based around the careful inspection of search results and data analyses, it is believed that an incremental approach can iteratively improve the integrity of the knowledge base.

During the review phase the data within the RKB was essentially frozen, but the results of queries and analyses performed have been cached. These results can now be used as a test base for the application of the co-reference analysis algorithms, which are to be deployed during the next phase of development.

Action: Continue research and development of co-reference technologies.

2. Navigation / Go back / Undo / Previous view

The RKB Explorer is a dynamic application, with parts of the display being updated asynchronously by javascript requests to the server. In addition, to avoid reloading the graph applet for each new item to be displayed, it too is dynamically updated. However, the down-side of this style of interaction is that the browser's forward and back buttons do not work in the usual manner, as the user has actually only requested a single initial page.

In an attempt to overcome this issue, the recently viewed menu was implemented to permit users to go back to a previous state. One user failed to find this facility at all. However, the majority of comments indicated that it was useful, but did not really replace the usual forward/back buttons. Furthermore, selecting a recently viewed item from the list presents a fresh graph containing only that resource, rather than the graph which the user created which potentially included other resources and/or a different layout.

Action: Investigate feasibility of various options for future RKB Explorer release. May include implementing back/undo button, saving graph states within the java applet, or adding options to toggle/disable dynamic behaviour.

3. Unclear how relationships are determined / linked in graph

Comments were received from some reviewers querying how relationships are identified, or why two resources were linked/related. This is often the case where resources are poorly defined or have little semantic information relating them to other resources, such as in the case of new students.

It is not intended for users to need to understand the detailed inner workings of the RKB Explorer application, and in any case these may be adjusted or evolve over time. Additionally, as co-references continue to be resolved, the analysis algorithms will become more accurate.

Action: Improve documentation/help text to include basic description of how relationships are derived.

4. Accented characters not displayed correctly

Several users noted that accented characters were not displayed properly, or appeared as HTML entities within the graph display.

Action: Resolved.

5. Errors or inconsistencies in source data

The RKB contains many tens of millions of facts, acquired from a variety of different sources. This quantity of data cannot be manually curated, and as a result errors and inconsistencies that exist among the original information sources are visible. In particular, the RKB's handling of names by one publications data source is less than ideal, with some records containing the names of two authors, or extraneous words such as 'and' or 'by'.

Action: Investigate feasibility of automated curation methods for common errors.

6. Detail panes could show more information

Users have suggested that more information should be presented by the RKB Explorer, or that certain data cannot be discovered. Indeed, the time available has not permitted the prototype to reflect as yet all information within the repository. The detail panes require further attention, showing additional data such as paper abstracts and links to full texts, where available, and details of affiliation to organisations.

Action: Complete implementation of detail panes.

7. Ordering of names and publications appears random

The ordering of results in the four columns in the lower half of the display is currently determined by the weighting or strength of relevance to the currently selected resource, or by date in the case of publications and projects. Several users enquired as to why the ordering of names appeared random, which in the case of people with the same weighting they were. Ordering is now additionally dependent on both weight and alphabetical sorting.

Action: Resolved.

8. Old browsers not supported

Older web browsers do not sufficiently implement Cascading Style Sheet and/or Javascript specifications to display the RKB Explorer correctly. Within the time and resources available, it is not possible to support legacy software that does not properly conform to web standards.

Action: Check documentation lists supported browsers. Investigate possible warning/notification for incompatible browsers.

9. Poor indication of activity when application is "busy"

On some browsers, there is little indication after a selecting a new resource that the system is busy. Code has been modified to show an hourglass cursor more prominently.

Action: Resolved.

10. Extraneous ontological entities appear

In some instances high-level ontological entities such as akt:Thing appear within lists of related items. These have been suppressed.

Action: Resolved.

11. Institutions are poorly represented

Some users expressed a wish to explore by institution/organisation, or to find our more details regarding institutions. Currently no information is presented regarding institutions as such, but it is intended that the detail pane for people should identify the organisations to which that person is affiliated.

Action: Add affiliations to detail pane for people. Consider feasibility of introducing an institution column in future release.

12. Confusion of use/action of top menu

Two reviewers expressed some confusion to the action and behaviour of the top menu. The top menu indicates the type of the current screen, whether it be person, research area, publication, project, or the search screen. Clicking on any of the non-selected types changes the display, to the most recently viewed resource of that type. If no resource of that type has been viewed, then a default resource is displayed. If the search tab is selected, then the result of the previous search (if there is one) is recalled, enabling a user to switch back to the search results after exploring a particular resource.

Action: Consider adding additional description to help/documentation

13. Missing associations with co-workers

One student enquired as to why they were not linked with their co-workers. No semantic information had been submitted by their site indicating affiliations to research groups or similar, hence there is no way of deducing such relationships. Where this information is present, it contributes a part in determining those who are considered relevant to an individual.

Action: None.

14. Distinguish between academics and students

One reviewer suggested visually distinguishing between academic staff and PhD students in people lists. While this will not be implemented, as lists are primarily ordered by closeness or importance to the currently selected resource, it is intended that the detail pane for people should convey these details.

Action: Revise detail pane for People to include status/job title.

15. Failed to find search

One user failed to find the search facilities. User-guide and help documentation appears sufficient; hence this has been attributed as user error.

Action: None.

16. Firewall issues prevented applet from loading

For one user, the RKB Explorer failed to load correctly. This has been attributed to an unusually restrictive corporate firewall, over which we have no control.

Action: None.

17. Improve results for scientific council members / reviewers

One user suggested improving results for scientific council members and external project reviewers. It is not intended that such results should be artificially skewed.

Action: None.

18. Liveness of imported data

One student expressed disappointment that a publication appearing in late 2006 had not been picked up by the RKB. In this particular case, the work was published more recently that the last capture of the ACM library into the RKB.

Action: Review acquisition from external data sources.

19. Names appear differently in list / graph / detail panel

It was noted that the name of an individual may differ in the representation presented within the lists, in the graph, and that of detail pane when the person is selected. These differences have occurred due to inconsistent retrieval of the label for a resource; however, code has since been updated to use the most appropriate label.

Action: Resolved.

20. ReSIST members and publications not distinguished

One user commented that it was not easy to distinguish between personnel and publications that are directly related to ReSIST, from those which are not. This is not an intended function of the RKB Explorer, which aims to cover the entire domain of resilient computing and dependable systems.

Action: None.

21. Searching with hyphens

One user encountered difficulties in searching for "*resilient-explicit computing*" –vs–"*resilient explicit computing*", ie omitting the hyphen. Currently the search text is passed to grep, enabling powerful regular expressions to be performed. However, it may be more appropriate to offer a more basic search as default, automatically replacing punctuation characters with wildcards.

Action: Review search mechanism. Amend documentation.

22. Selecting more than one resource / union of selection

One user discovered that multiple resources within the graph display could be selected at any one time. They reasonably assumed that the display would update to reflect the union or overlap between the two resources, but this (interesting) functionality has not been implemented. Unfortunately time constraints will not allow for this development.

Action: None.

23. Search for surname/forename -vs- forename/surname

One user commented that searching for "BLOGGS, Joe" did not return any results, whereas searching for "Joe Bloggs" did. All names are stored within the repository as "given-name family-name", and searches for people should be entered in this fashion.

Action: Review help/user guide documentation.

24. Tooltips on graph controls

It was suggested that tooltips were added to the controls within the graph display. In fact such tooltips are already present, but in some browsers they are not displayed on mouse-over unless the Java applet currently has focus within the browser. This behaviour is dependent on the browser, and cannot be overcome.

Action: None.

25. Unclear of which resource/entity is the current focus

One user reported difficulty in identifying which resource the display is currently reflecting. This is clearly displayed in the title bar of the graph view, and in the accompanying detail pane. It is also explained within the help/documentation. Attributed to user error.

Action: None.

26. Confusion over unknown/withheld persons

One user queried why there are persons named “Unknown/withheld”. These entities arise when there is a missing name attribute, or in the case where an institution has exported a relationship identifying a person by URI, but withheld details of that person due to data protection or other legal action.

Action: Consider adding information to user guide / help documentation.

27. Pagination of search results

For performance reasons, only the first 100 matching literal values are processed when performing a search. Options should be made available to the user to retrieve the “next” 100 search results, if the desired resource is not found within the first set.

Action: Implement proper pagination of search results. Consider reducing number of results retrieved per ‘page’.

28. Confusion over search results layout

One user appeared to be confused as to the format of the search results. On querying for a literal value matching resources are placed in the appropriate column, depending on whether it is a person, research area, publication or topic. It is believed that in the case in question, the user queried their own name, and was presented with a single (well co-referenced) name in the person column, and a single paper with their name in the title in the publications column, whereas they appear to have been expecting related papers and projects related in those columns.

Action: Review user guide / help documentation.

Conclusion

The review process has, on the whole, validated the RKB Explorer interface as a useful tool, which has been well received by the users. The development team already knew many of the issues raised; however as a result of this process several other important points have been revealed and resolved.

As listed in the above section, there are several actions to be implemented as a matter of course throughout the continued maintenance and minor development of the RKB and the Explorer interface.