

Project Specification — Information Aggregation

David Osbourn

May 25, 2005

Organisation

British Telecommunications plc.
Enterprise House
84 - 85 Adam Street
Cardiff
CF24 2XF

Outline

Information aggregation is the bringing together of different information sources. Such information sources are considered to be information assets, whether they be documents, images or multimedia data. What is important is that when information is aggregated it is often possible to derive and imply information which is not explicitly included in any of the data sources. It is possible that whilst the individual information assets contain information considered to have a low sensitivity level, the derived, aggregated information may have a higher sensitivity level.

A very simple example of information aggregation could be two press releases from an organisation. The first statement might include the organisation's annual turnover, while the second statement could include their annual costs. It is very simple to then derive the organisation's annual profits from these two statements.

In this simple example the information assets were put out in to the public domain, i.e. the organisation has deemed this information to have a low sensitivity level. The implication of this is that information with a higher sensitivity level is also now in the public domain. Perhaps the mistake was to allow both turnover and costs to be present in two similar information assets.

Clearly there are two issues here: firstly, how are the individual information assets controlled from both a content and access point of view? Secondly, how can the aggregated information be controlled?. In other words, how is it decided what data to include in each information asset, and then how can access to that information asset be controlled, such that the aggregated information is also controlled?

There are numerous other examples which provide much more complicated and in depth case studies with regards to information aggregation. For example, how can BT control access to maps of their communications networks, where highly sensitive information about government and national infrastructure is present, and derivable from certain parts of the map? Similar examples exist within other areas of the critical national infrastructure, e.g. water supply, communications or power networks.

Benefits

The proposed project is not currently related to my day-to-day work role. As such it is difficult to identify any particular benefits to BT from this work. How-

ever, the obvious benefit is to obtain an increased understanding of information aggregation, as well as a set of guidelines that may benefit individuals also faced with this problem.

This is a real business issue, raised to me by my industrial sponsor and one that has been confirmed as a worthy area of research by the *National Infrastructure Security Co-ordination Centre* (NISCC)¹ and also by two individuals working for *QinetiQ*².

Deliverables & Project Aims

The aim of the project is to investigate scenarios similar to the example listed above. In particular the deliverables of the project will be:

- A number of prevalent examples which highlight the problem in a business context.
- A definition of the problem using tools such as *soft systems methodology*, to analyse the provided examples. Justification will be provided for the choice of tool.
- An overview and critical analysis of existing work, including the solutions to related problems, and reasons whether they do, or do not, provide a suitable approach.
- A set of recommendations which outline a best practice solution to the identified problem.
- An abstract model to accompany and support implementation of the recommendations.
- An appraisal of the level of success of the project.

In addition to these it is also an aim of the project to deliver a prototype software component which implements the recommendations and model detailed above. This though is dependent on the completeness of the solution and suitable time being available after the key deliverables are produced. Reasons for either including or omitting this will be detailed in the review of the success of the project.

The written portions of the deliverables will be presented in a dissertation report in accordance with academic standards set out by the University of Cardiff and the university's department of Computer Science.

¹<http://www.niscc.gov.uk/niscc/index-en.html>

²<http://www.qinetiq.com/>

Constraints & Quality Plan

The main constraints to consider at this stage are:

- The project and deliverables are to be completed and submitted to the university for marking, at the latest, by February 28, 2006.
- The names of any BT clients, or any information covered by BT's corporate security policy should not be included in the project dissertation.
- It is not intended that any of BT's time will be used to fulfil the anticipated workload. If this becomes untenable, then the option of arranging study leave will need to be broached.
- At this early stage it is still not possible to state which particular software tools will be used. My personal preference is to use Open Source/GPL software and not to use Windows based applications. This decision will not have any detrimental effects on the delivery or quality of the project.

Project Plan & Timescales

A detailed project plan has been developed independently of this project specification. The plan lists, at a detailed level, the tasks that make up the project work. The plan also provides a detailed Gantt Chart which lists the projected timescales associated with each task. The Gantt chart also identifies the milestones within the project.

Academic Sponsor

Jeremy Hilton

Computer Science,
Cardiff University,
Queen's Buildings,
The Parade,
Cardiff
CF24 3AA

Jeremy.Hilton@cs.cardiff.ac.uk

Industry Sponsor (Informal)

Jonathan Powell

ppE1A, Enterprise House,
84 - 85 Adam Street,
Cardiff
CF24 2XF

Jonathan.Powell@bt.com