

Two Case Studies for KAKTUS

Mark Burnett

DSTO C3 Research Centre
Fernhill Park
Department of Defence
Canberra ACT 2616
AUSTRALIA

Mark.Burnett@dsto.defence.gov.au

Abstract

KAKTUS is a step toward what Tim Berners-Lee, the creator of the World Wide Web, calls a "semantic Web" where people, software agents, search engines and other programs can gain access to meaning—rather than just content—on a Web site. A semantic Web potentially also lets programs utilize all the data on Web pages, allowing them to gain knowledge from one site and apply it to logical mappings on other sites.

Keywords Semantic Web, Enterprise model, Textual recognition.

1 A handle on context

In a world where information environments are becoming larger, more complex and more dynamic, KAKTUS uniquely provides a contextual view of information. KAKTUS recognises that the problem of identifying meaning within text is dependent on the context within which the text is read. It effectively allows different people to have different models of the same data. And by representing the model in an open and transformable way, KAKTUS allows those views, or islands of knowledge, to be shared and exploited within the enterprise.

2 Building a semantic web

KAKTUS builds a semantic web by automatically populating a model of an enterprise, scenario or organisation with data taken from documents. Pattern matching and basic natural language understanding techniques are used for textual object recognition and identification, and provide a low-cost furnishing of the information environment for KAKTUS. These techniques also allow for relatively easy creation, adaption and maintenance of the webs by a trained user or external consultant.

3 Business objects and processes

KAKTUS provides a framework for users to view information through the lens of business objects such as people, places and things and business processes that relate those objects within an enterprise scenario. This is effectively a gearing process that models information at a level that is meaningful to a user, above the unit level of documents or chunks of text.

4 Capabilities

A number of different capabilities or applications can be surfaced from the semantic web. There are applications for information retrieval, information visualisation, knowledge management and fact extraction. The following is an indicative list of possibilities that have already been prototyped.

- Advanced search capabilities that surfaces information based on semantic relatedness to a query term;
- A key list of terms related to a user query based on physical (within-document) and semantic proximity. This can include association of people (as objects within the semantic model) with information;
- Summary of information with respect to a particular topic across a web site or document collection that gives coverage of both the semantic and document dimensions;
- Dynamically hyperlinked web sites, with the hyperlinks relevant to the context and content of the user's information need;
- Extraction of facts - taken here to mean data with a fixed semantic correlation - or partial facts from within a collection of texts.

5 Features

KAKTUS builds a framework for managing information and knowledge with the following attractive end-user features:

- Provides value from day 1 - it doesn't require a "big bang" population of a huge model;
- Scalable - the underlying techniques scale in all directions;
- Flexible - a semantic web can service an entire enterprise over an extended period, or a single analyst for one day;
- Complimentary to existing tools such as search engines

6 Demonstration

Two cases studies for KAKTUS will be demonstrated, illustrating how the approach can be used in diverse scenarios. The first involves planning documents for DSTO for the year 99-00. In this case there's a complex underlying model of the enterprise involving DSTO, Defence, commercial organisations and their inter-relation.

The second involves PROMED emails. PROMED, the Program to Monitor Emerging Diseases, is the premier discussion group for specialists in emerging infectious diseases. It was proposed by the Federation of American Scientists (FAS) to create a global system of early detection and timely response to disease outbreaks. A relatively simple semantic model underlies the reporting of disease and disease outbreaks for PROMED.

For both case studies we demonstrate: (i) association of key related information following a user query against the document set, based on semantic and physical (within document) proximity; (ii) a collection wide summary of information related to a user-defined topic, that takes account of both key topic coverage and document coverage; (iii) navigation of the information based on dynamically hyperlinked HTML, with links generated in response to the KAKTUS-derived set of related topics.

For PROMED we also show how an automated approach to fact or partial-fact identification within the emails.