

**DP0343042: Investigators: Eklund, Cole, Bruza and Song. 2003/2004/2005****Title: Concept learning and Structure Formation for Document Navigation — 2003 Report**

The aims of the project are to develop (i) a framework motivated by a combination of Hyper-index analogue to language and formal concept analysis (FCA) (ii) to develop context sensitive conceptual spaces with which to browse and retrieve documents.

The development of a theoretical and experimental framework in which to test the main hypothesis of the work, that conceptual structures and scales can be automatically inferred from document content. This necessarily involves an understanding and summary of the relevant literature. In particular much effort has been in the APD understanding the Information retrieval aspects of the project which are introduced by the skills of Bruza and Song and incorporating these with knowledge of FCA held by Cole and Eklund.

The year was punctuated by 4 conf papers corresponding to 5 variations on the theme of deriving logical structures using FCA. Work presented at the 1st conf, ICFCA'03 related to a space of taxonomies organising previous work combining FCA and software engineering. This work used information such as authors, citations, and a hand crafted controlled vocabulary to construct a system of concept lattice organising the paper. ICFCA'03 coincided with a 2 week visit with Rudolf Wille's group in Darmstadt, the headquarters for FCA. Work presented at the 2nd conf, SEKE'03 held in June, related to context sensitive navigation within a wide variety of taxonomical structures derived automatically from program source code. In the 3rd conf, ICCS2003, Eklund presented a summary of the global research effort in an invited lecture. The 4th conf, ADCS 03, related to experiences navigating a text collection with a tools whose prime mode of display is a term-document sub-lattice. The last conf publication for ICFCA'04 (yet to occur) relates to a technique to combine and contrast information from different observers who make different judgments about a situation.

The common theme drawing these publications together is the use of FCA to derive a space of taxonomic structures in which the user can navigate in order to learn about the structure and retrieve items and facts in a goal directed manner. A further paper in ICFCA04 describes experiments with usability that are important to gauge the relative success or otherwise that results when information users are asked to navigate documents using concept lattices.

Much time during the year has been devoted to work (yet to be reported formally). This work focuses on text processing experiments involving clustering both documents and terms using a probabilistic bi-gram model induced by the HAL matrix. The work has involved the construction of software programs and a review of literature related to the topic. Important within this topic is the notion that the relevance of a document to a query involves the user. As a user becomes more familiar with nature of a document collection in which they are searching they become better able to describe their information need(s). Thus it is important for the system, in a user specific manner, to be able to communicate information that will help the user better formulate their query.

**Applications:** Within information system scalability is an important issue. To this effort has been devoted to the construction of scalable software systems. Work has proceeded on the following open source projects: Set and Relations Library (SARL) provides the basic data structures required in FCA: namely sets, relations, formal contexts, lattices, partial orderings, order embeddings, covering relations, and lattice products. KB (Knowledge Base) implements a very fast memory mapped triple and a graph based query language used to construct concept lattice organising aspects of the content of the knowledge base. Tockit - Treatment of Conceptual Knowledge In Tockit is a collaborative effort to construct software tools around the theory of FCA. The plan is too consolidate the text processing experiments already constructed to the point where the results are suitable for publication in conf proceedings. Larger and more systematic usability testing is also planned.

**Publications:** 1 edited book[B10.2(2)], 1 journal article[B10.2(6)], 1 book chapter[B10.2(3)] 4 refered conf. papers[B10.2(12)] + 3 papers by R. Cole], 1 provisional patent[B10.2(37)].

**SR0354513: Investigators: Dale et. al, The Computational Processing of Human Language. 2003**

Preliminary coordination meeting and web-site has been established. Network is being expanded to International contacts and consolidated with other like networks.

**SR0354476: Investigators: Nayak et. al, Intelligent Applications Through The Semantic Web. 2003**

Preliminary coordination meeting and web-site has been established. Network is being expanded to International contacts and consolidated with other like networks.