

Company Overview

Website: www.c-content.nl

Established in 1987, our organization is one of the oldest and most experienced suppliers of search engine technology, content publishing and information retrieval technology for professional users worldwide.

Our technology aims at easy and efficient access of large, multi-lingual, mission-critical databases for a variety of users, and our search and retrieval solutions are particularly suitable for official, legal and scientific publications. The main characteristics of our products are the extremely fast search speed, reliability under heavy load conditions, multi-linguality and flexibility in interface design.

During the past 6 years, we have mainly concentrated on development of content integration solutions for specific market segments, such as legal professionals.

As a result of this development strategy, in May 2006, we launched a product specifically designed for content integration in the legal market. This legal content integration system offers immediate and appropriate access to all relevant legal information in the Netherlands, by means of an intuitive user-friendly interface. This application provides integral access to three types of sources:

- External legal sources (of all the legal publishers), based on subscription.
- Public legal sources.
- In-company documents and document management systems (optional).

In addition, we continue development of the content integration system towards a universal content integration system. Such a system can easily be adopted to support all relevant information in a specific (professional) domain accessible via one easy to use search system. The main characteristics of this universal system are:

- Simultaneous access to various on and off-line sources ("Federated Search").
- Daily update of documents from all sources
- Automatic classification of all documents based on pre-defined classification structure.
- All relevant documents are connected with each other.
- The most relevant documents are presented on top.
- Extensive possibilities to narrow-down the search-results (drill-down).
- Immediate access to the proper documents, by means of the very fast search engine.

Plus a range of user-related functionality, such as:

- The possibility to mark relevant text fragments, save and share them with colleagues as know-how documents.
- Thesaurus suggestions to broaden or narrow-down the search-results.
- Alert mails to inform a user when new, relevant information becomes available.

The response from the market is extremely positive. In our opinion, content integration fulfills the user's ultimate desire: access to all relevant (professional) information through a single and easy to use search system. As such, content integration is generally considered as the future of information retrieval.

Graduation assignment

Personalization based on user behaviour

Research the personalization options present in the field of information retrieval. Which personalization options occur frequently and which are the latest techniques? To what extent are these options available in the legal content integration system and which options would strongly improve the use of the system for its end-users? What upcoming techniques do you foresee and to what extent are these relevant for the improvement of the user-experience of the legal content integration system? Which personalization options will be further explored and realized will be based on this analysis.

Seminar Assignment:

Combined ranking of federated search

How can the results of two (or more) search engines be merged and ranked by relevancy, given that each search engine uses a different ranking method. The first goal is ranking of the federated search results of the following search engines:

- Domain-specific system, where a number of parameters influence the search results, like date, type of information, adjacency, and standard 'best match' indicators
- System based on elaborate TF-IDF ranking

The ranking scores of both systems are available in the federated search results, and a normalization mechanism is required to provide a combined search result ranked by relevancy.