**Проекти по “Изкуствен Интелект” за спец. “Компютърни науки” 2009-2010**

# Smartbook

1. Да се направи търсене в книги по, като се използва TF/IDF(Lucene).
2. Да се направи търсене на книги, като освен TF/IDF(Lucene) се използва и профил на потребителя(ключовите думи извлечени от предишния проект).
3. Да се направи търсене на книги, като освен TF/IDF(Lucene) се използва и онтология която ни помага за семантично търсене.
4. Правене на разюме на книга (summarization)
5. Правене на разюме на книга въз основа на заявка за търсене от потребителя (query based sumarization)
6. Правене на разюме на книга (въз основа на ключови думи и семантика) – да се ползват topic maps.
7. Да се модифицира горния алгоритъм за да може да работи с променящи се данни и да се използва за препоръки на нови продукти на потребителя.
8. Автоматично анотиране на текст – да се види системата CLARK ([www.bultreebank.org](http://www.bultreebank.org/))
9. Да се използва fedora(<http://www.fedora-commons.org/>) за сървър за съхранение и управление на книги, като на него се интегрират търсенията обяснени по горе.
10. Да се анотира текст като се използват услуги от <http://www.clarin.eu/>
11. Да се анотира текст като се използват услуги от <http://www.opencalais.com/>
12. Автоматичен четец на книги на Български (английски)
13. Анотация PST на Български (английски)
14. Chatbot – проучване на програмите направени до момента, анализ на проблемите, които се срещат в съществуващите решения и предлагане на ваше решение. Реализиране на решението. (1-3 човека според зависи от широчината на изследването и постигнатите резултати.)

INEX 2010 Book Track

### Overview

The goal of the Book Track is to promote inter-disciplinary research investigating techniques for supporting users in reading, searching, and navigating the full texts of digitized books and to provide a forum for the exchange of research ideas and contributions. Focusing on topics of interest in the fields of information retrieval (IR), human computer interaction (HCI), digital libraries (DL), and eBooks, the track in 2010 will explore the following four tasks:

* **'Prove It' search task:** Find evidence in books to confirm or reject a factual statement. Systems need to return paragraph XML elements or passages from books that contain evidence regarding the statement, either confirming it or rejecting it. This task tests the application of focused retrieval approaches to a collection of over 50,000 digitized books to return users relevant book parts.
* **'Best Books to Reference' search task:** Find the best, most relevant books on the subject of a given factual statement. This task tests domain-specific full-text search methods on a collection of over 50,000 digitized books.
* **Active Reading task**: Conducting user studies into active reading, i.e., exploring how and why readers use eBooks in specific scenarios with a focus on eBook usability.
* **Structure Extraction task**: Building navigation tools for digitized books by constructing hyperlinked table of contents from OCR text and layout information for a sample of 1,000 books.

### Book Corpus

The track builds on a collection of over 50,000 digitized, out-of-copyright books, provided by Microsoft Live Book Search and the Internet Archive (for non-commercial purposes only). The OCR content of the books is stored in an XML format, referred to as BookML. Most books also have an associated metadata file (\*.mrc), which contains publication (author, title, etc.) and classification information in MAchine-Readable Cataloging (MARC) record format.

For instructions on how to access the book corpus, please refer to the website: <http://www.inex.otago.ac.nz/tracks/books/books.asp>.

**Participation**

Participants are expected to contribute to at least one of the tasks listed above, as well as to the construction of the book test collection. The latter involves the creation of test topics (factual statements that can be found in a book from the corpus) and the contribution of relevance assessments. Participants will gain access to the constructed test collection.

# DataMining/Machine Learning

1. Разработка на система за DataMining с SOA архитектура:
   1. Разработка на модул за крос валидация
   2. Разработка на T-pair test
   3. Разработка на Wilcoxon
   4. Разработка на графичен интерфейс
   5. Сравнение със weka като бързодействие и резултати!!
   6. Разработка на алгоритми
      1. Основни алгоритми за клъстеризация и класификация.
      2. Клъстеризация на голям обем от данни.
      3. Класификация на голям обем от данни.
2. Разработка, анализ и модификация на алгоритми за Машинно Самообучение с прототипи в системата Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) . Проекта е за 1-3 човека.

# Game framework

1. Да се разработи сайт за интернет игри, като за всяка игра се реализира ИИ.(всички игри трябва да използват една технология например Java Aplet)
   1. Разработване на логически и математически игри(очаквам идеи и предложения в тази точка.)