

CS491-492 Senior Project Topics

FAZLI CAN
Computer Engineering Department
Bilkent University
Bilkent, Ankara 06800

September 7, 2007

I provide a tentative description of the projects that I plan to supervise this academic year. Their description is in progress.

I am looking for enthusiastic students with imagination and who are good at four core practices [innovating, modeling, programming, and systems thinking]¹ of computer science. So far I have two tentative groups of students.

1. **Web Crawler Project:** The purpose of this project is to build a web crawler (http://en.wikipedia.org/wiki/Web_crawler) which visits Turkish newspaper web sites, and crawl them for daily news. The parser should modify itself since Web news resources frequently change their Web page (news reporting) structure. If the parser is dependent to the html tags or lines on a Web page it can be obsolete within a short amount of time. However, if it handles text by checking the population of tags on a page, it can be used at any news page. The project will go in steps: first crawlers for specific sites will be developed then a general purpose crawler will be implemented.
2. **Summarization Project:** Text summarization is important in several applications such as information retrieval (IR) and new event detection and tracking (NEDT). In IR rather than seeing the contents of the retrieved documents users may prefer to see the summary of individual documents. In NEDT it is possible to have single document summarization or multi document summarization. In such environments it is also possible to generate summaries according to the user interests (e.g., query oriented summary). The purpose of this project is to generate a standard test collection for Turkish summarization and develop algorithms and applications for a news portal. One possible sub component of this project is snippet generation.
3. **Style investigation of Ottoman poetry (divan siiri):** In this project we plan to investigate the stylometric elements and change in Ottoman poetry with time using text mining techniques. For this project you do not need to know the old script (eski yazi) or Ottoman poetry meters (such " failatün failatün failün"). In the project

¹ Denning, P. J, McGettrick, A. "Recentering computer science." *Comm. of the ACM*. Vol. 48, No. 11, 2005, pp. 15-19.

transcribed texts (letters in latin alphabet with some special symbols) will be used. Some previous studies on Turkish: Style Change With Time:

<http://www.users.muohio.edu/canf/papers/chum04.PDF>,

Ince Memed

short version: <http://www.users.muohio.edu/canf/papers/CHUM355Apr0504.pdf>,

long version: <http://www.users.muohio.edu/canf/papers/imstyle122504.pdf>

4. *Üç İstanbul Project*: This one is for people who are interested in the Turkish literature. It will be on generating the hypertext version of Mithat Cemal Kuntay's *Üç İstanbul* and discovering the patterns on the novel parts and characters (Adnan, Süheyla, Belkıs, Hüsrev and others). We also plan to look at the correspondence and relationship between the TV series and text versions. For inspiration please see the hypertext version of Faulkner's *The Sound and The Fury*:

<http://www.usask.ca/english/faulkner/>