

CS6362.501 Software Architecture and Design
Spring 2003
Project 2: A Web Search Engine
Due: 2:00am Monday March 17

1. Project Summary

As software architects, your team is to architect a web search engine, Microminer, using KWIC⁺ which is an extension to the KWIC* software system you have designed and implemented in Project 1. For this project, you will continue using the architectural style you've chosen in the first project.

2. Microminer - A web search engine

Functional requirements: Microminer shall accept a list of keywords and return a list of URLs prefixed with their descriptions that contain any of the given keywords. No noise word shall be given as part of the list of keywords. Microminer shall use another software system, KWIC⁺ - a KWIC (Key Word In Context) index system, in order to efficiently store the URLs and the corresponding descriptions. KWIC⁺ shall accept an ordered set of lines, where each line consists of two parts: the descriptor part and the URL part:

- The URL part, whose syntax is:
URL ::= 'http:/' {identifier '.'}+ ['edu' | 'com' | 'org' | 'net']
identifier ::= {letter|digit}+
letter ::= ['a' | 'b' | 'c' | ... | 'y' | 'z' | 'A' | 'B' | ... | 'Y' | 'Z']
digit ::= ['1' | '2' | '3' | ... | '9' | '0']
- The descriptor part, whose syntax is:
{letter⁺}⁺.

The descriptor part of any line should be *circularly shifted* by repeatedly removing the first word and appending it at the end of the line. The KWIC⁺ index system shall output a list of all circular shifts of all lines in ascending alphabetical order ' <a<A<b<B<...<y<Y<z<Z, together with their corresponding URLs. No line in the output list shall start with any noise word such as "a", "an", "the", "and", "or", "of", "to", "be", "is", "in", "out", "by", "as", "at", "off".

Non-functional requirements: The Microminer system shall be easily understandable, portable, enhanceable and reusable with good performance. The Microminer system must also be user-friendly, responsive, and adaptive.

3. The Deliverable

Your descriptions should be elegant and comprehensible. Your deliverable should be available as specifications:

- **The process architecture – management and organization of your team work:** describe how your team members were divided up in carrying out your own design tasks, and why each took the particular roles. In other words, describe the essential tasks as components of your own process architecture, their essential relationships as interactions, while taking into consideration other architectural concerns. In addition, describe how your team was organized and communicated with each other, e.g., hold meeting regularly, work closely together, or mostly independent, etc.
- **Architecture specification:** describe both pictorially and textually, architectural style, components/elements, interactions/connections, any constrains, any patterns, and a discussion of advantages and disadvantages of your chosen architecture. Describe how to add two new functions, deleting noise words and search, with at least two options for each function in your

architecture. Give a discussion of advantages and disadvantages of each option of each function. Based on this analysis, choose one option for each function for your system and explain the reasons

- **Java applet specification:** build an applet that should run with a web browser. Your program should be well documented and tested.
- **User manual:** describe how the user can access and use the system. Your descriptions should describe the typical interactions between the user and the system, e.g., what are the steps the user has to follow in using the system. Use screenshots, if needed, to show how the system looks like initially as well as for subsequent steps that the user takes.

4. Project Submission Policy

Projects must be submitted to WebCT before the deadline, which is 2:00am on March 17. It is your responsibility to correctly submit the project to WebCT. A penalty of 30% will be deducted from your score for the first 24-hour period your project is late. A penalty of 70% will be deducted from your score for \geq 24-hour period. No credit for \geq 2 days. Weekend days will be counted.