

## CS 533 Homework #1 - Group 4

### Natural Language Processing for Information Retrieval (IR)

by David D. Lewis and Karen Sparck Jones



IR is for **the user who wants to learn something**, as opposed to merely wanting a specific data item or question answered.

IR must find **relationships between the information needs of users and the information in the documents**. This relationship between the user's need and the text that meets it is **not necessarily obvious**.

IR depends on **indexing for retrieving documents relevant to the user's need**. The aim of indexing is to **increase precision and recall**. However, users have little skill or limited experience in formulating initial search requests, thus requests are *underspecified* and *incomplete*.

Then, the question is **whether information retrieval should be using Natural Language Processing (NLP) or not**. Possible applications of NLP in IR: **text processing, indexing, searching and matching**.

The **challenges** from NLP point of view: **whether NLP is doable** and **whether non-statistical and statistical data can be appropriately combined**.

There is strong requirement for **sophisticated NLP-based approaches to indexing and retrieval**.

**Major challenges** are making the technology operate **efficiently** and **effectively** on the appropriate **scale**, and conducting the **evaluation** tests essential to determining whether the approach works especially when the tests involve interactive searching of **large files**.