



# Basics of Information Retrieval

---

Lillian N. Cassel

Some of these slides are taken or adapted from

Source: <http://www.stanford.edu/class/cs276/cs276-2006-syllabus.html>

# Basic ideas

- ❑ Information overload
  - The challenging byproduct of the information age
  - Huge amounts of information available -- how to find what you need when you need it
    - Think about addresses, e-mail messages, files of interesting articles, etc.
- ❑ Information retrieval is the formal study of efficient and effective ways to extract the right bit of information from a collection.
  - The web is a special case, as we will discuss.



# Some distinctions

- ❑ Data, information, knowledge
  - How do you distinguish among them?
    - <http://www.systems-thinking.org/dikw/dikw.htm>
- ❑ Information sources
  - Very well organized, indexed, controlled
  - Totally unorganized, uncharacterized, uncontrolled
  - Something in between