Course Overview

Kuan-Yu Chen (陳冠宇)
2019/09/20 @ TR-514, NTUST
Information Retrieval

- Information Retrieval is a broad area of computer science
  - It mainly focuses on proving the users with easy access to information of their interest

- IR deals with the representation, storage, organization of, and access to information items
  - documents, Web pages, online catalogs, structured records, multimedia objects
The Goal of IR.

• Early goals of the IR area
  – Indexing text
  – Searching for useful documents in a collection

• Nowadays, research in IR
  – Modeling
  – Web search
  – Text classification
  – Systems architecture
  – User interfaces
  – Data visualization
  – Filtering
  – Languages
The Goal of IR...

- In terms of research, the area may be classified into two distinct and complementary points
  - Computer-centered
    - Building efficient indexes/representations
    - Processing user queries with high performance
    - Developing ranking algorithms
  - Human-centered
    - Studying the behavior of the user
    - Understanding the information need
General Architecture of the IR System

Crawling Process

Inverted Index

Indexing Process

Retrieval & Ranking Processes

Answer Set:
1. document d_{i0}
2. document d_{i23}
3. document d_{i5}
4. document d_{i47}

User query

Query parsing & expansion

System query

Retrieval & Ranking Processes
IR Problems.

• Users of modern IR systems, such as search engine users, have information needs of varying complexity
  – A full description of the user information need is not a good query to be submitted to the IR system
  – Instead, the user translates this information need into a query
    • A set of *keywords*, or *index terms*, which summarize the user information need
  – The key goal of the IR system is to retrieve information that is useful or relevant to the user’s information need
IR Problems..

• That is, the IR system must rank the information items according to a degree of relevance to the user query.

• The definition of the IR problem
  – *The key goal of an IR system is to retrieve all the items that are relevant to a user query, while retrieving as few nonrelevant items as possible.*

• The notion of *relevance* is of central importance in IR.
About “Relevance”

- Relevance is a personal assessment that depends on the task being solved and its context.

- Relevance can change with:
  - Time
  - Location
  - Device
  - ....

Until now, no IR system can provide perfect answers to all users all the time!
Tentative Grading

- Homework: 75%
  - Programming with C/C++/Python/Matlab
    - HW1: 15%
    - HW2: 15%
    - HW3: 15%
    - HW4: 15%
    - HW5: 15%

- Attending the Invited Speech: 10%

- Challenge: 20%
  - Presentation: 10%
  - Competition: 10%
<table>
<thead>
<tr>
<th>Date</th>
<th>Syllabus</th>
<th>Homework</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/13</td>
<td>Holiday for Moon festival</td>
<td></td>
</tr>
<tr>
<td>9/20</td>
<td>Course Overview</td>
<td></td>
</tr>
<tr>
<td>9/27</td>
<td>Classic Models</td>
<td>Homework-1 (deadline: 10/9 23:59)</td>
</tr>
<tr>
<td>10/4</td>
<td>Extended Probabilistic Models</td>
<td>Homework-2 (deadline: 10/17 23:59)</td>
</tr>
<tr>
<td>10/5</td>
<td>Break for RoCling 2019</td>
<td></td>
</tr>
<tr>
<td>10/11</td>
<td>Holiday for National Day</td>
<td></td>
</tr>
<tr>
<td>10/18</td>
<td>Evaluation &amp; Benchmark Collections</td>
<td></td>
</tr>
<tr>
<td>11/1</td>
<td>Search Results Diversification</td>
<td></td>
</tr>
<tr>
<td>11/15</td>
<td>Speech &amp; NLP (Mr. Hsiao-Tsung Hung, ASUS)</td>
<td></td>
</tr>
<tr>
<td>11/22</td>
<td>Representation Learning for Information Retrieval</td>
<td>Homework-5 (deadline: 12/5 23:59) &amp; Submit Your Member List!</td>
</tr>
<tr>
<td>11/29</td>
<td>Supervised Retrieval Models &amp; Web Search</td>
<td>Submit Your Paper Title!</td>
</tr>
<tr>
<td>12/6</td>
<td>Final Competition (deadline: 12/20 8:00)</td>
<td></td>
</tr>
<tr>
<td>12/13</td>
<td>Break for ASRU 2019</td>
<td></td>
</tr>
<tr>
<td>12/20</td>
<td>Presentation 1</td>
<td></td>
</tr>
<tr>
<td>12/27</td>
<td>Presentation 2</td>
<td></td>
</tr>
<tr>
<td>1/3</td>
<td>Presentation 3</td>
<td></td>
</tr>
</tbody>
</table>
Instructor & TA

- Instructor: 陳冠宇
  - E1-222-4, Thursday 9:00~16:00
  - kychen@mail.ntust.edu.tw
  - (02) 2737-6377

- TA: 徐子杰 (E1-222-3)
References.


References

• Conferences
  – ACM Annual International Conference on Research and Development in Information Retrieval (SIGIR)
  – International Joint Conferences on Artificial Intelligence (IJCAI)
  – ACM Conference on Information Knowledge Management (CIKM)
  – Annual Meeting of the Association for Computational Linguistics (ACL)
  – International Conference on Learning Representations (ICLR)

• Journals
  – Journal of the American Society for Information Science (JASIS)
  – ACM Transactions on Information Systems (TOIS)
  – Information Processing and Management (IP&M)
  – ACM Transactions on Asian Language Information Processing (TALIP)
  – Information Retrieval Journal (IRJ)
Questions?

kychen@mail.ntust.edu.tw