

Web工学 (Web Engineering)

<http://tinyurl.com/webeng2015>

豊田正史 (Masashi Toyoda)

E-mail: mtoyoda@acm.org

Goal

- Learn basics of Web related research area and its recent trends
 - Search engine, Information Retrieval
 - Natural Language Processing
 - Machine Learning, Data Mining
 - Social Media/Network
 - Big Data
 - Cyber-Physical Systems
 - Internet of Things (IoT)
 - etc. etc...

Format

- Seminar Style
- Three or four students present research papers (published on top conferences/journals) at each session
 - Presentation: 15 min.
 - Q&A: 10 min.

Organization of sessions

- First half: Basic papers from 1998-2007
 - Google PageRank, HITS,...
- Second half: Recent research papers
 - WWW, WSDM
 - SIGMOD, VLDB, ICDE
 - SIGIR
 - KDD, ICDM
 - AAAI, ICWSM
 - ACL, EMNLP
 - Etc.

Credits

- Point system:
 - 100 points (優)
 - 80 points (良)
 - 60 points (可)
 - Less than 60 points (不可)
- Scores:
 - Presentation: 60 points
 - Good question/comment: 20 points
 - Attendance: 0 points

Schedule

- 10/ 5 (Mon) Guidance
- 10/12 (Mon) Holiday
- 10/16 (Fri) Deadline: Submit a list of papers to read
- 10/19(Mon) Session
- 10/26(Mon) Session
- 10/29(Thu) Session
- 11/ 2(Mon) Cancel
- 11/ 9(Mon) Session
- 11/16(Mon) Session
- 11/23(Mon) Holiday
- 11/30(Mon) Session
- 12/ 7(Mon) Session
- 12/14(Mon) Session
- 12/21(Mon) Session
- 12/23(Wed) Session(Holiday)
- 1/ 7(Thu) Session
- 1/11(Mon) Session

Papers for Session 1

- Search engine and PageRank (担当 : 城戸祐亮)
 - Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW7, 1998. (The Seoul Test of Time Award. 2015)
- Crawler (担当 : 小林賢)
 - A. Heydon and M. Najork. Mercator: A scalable, extensible web crawler. WWW8, 1999.

Papers for Session 2

- Mirror detection (担当：依田伸樹)
 - Andrei Z. Broder, Steven C. Glassman, Mark S. Manasse, and Geoffrey Zweig. 1997. Syntactic clustering of the Web. In *the sixth international conference on World Wide Web*, 1997
 - Web Search (担当：三浦信一)
 - Andrei Broder. 2002. A taxonomy of web search. *SIGIR Forum* 36, 2, 3-10, 2002.
 - Inverted list compression (担当：池下克彦)
 - M. Zukowski, S. Heman, N. Nes, and P. Boncz. Super-scalar RAM-CPU cache compression. In Proc. of the Int. Conf. on Data Engineering, 2006.
- OR
- V. Anh and A. Moffat. Index compression using fixed binary codewords. In Proc. of the 15th Int. Australasian Database Conference, pages 61–67, 2004

Papers for Session 3

- Link Analysis
 - Page, Lawrence; Brin, Sergey; Motwani, Rajeev; Winograd, Terry. The PageRank Citation Ranking: Bringing Order to the Web. 1999.
 - J. Kleinberg. Authoritative Sources in a Hyperlinked Environment, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46(1999). Also appears as IBM Research Report RJ 10076, May 1997.
 - Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins. Trawling the Web for Emerging Cyber-Communities. WWW8, 1999

Papers for Session 4

- Structure and evolution of the Web
 - Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener. Graph structure in the web. WWW9, 2000.
 - Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener. A Large-Scale Study of the Evolution of Web Pages. WWW2003, 2003.
 - A. Clauset, C.R. Shalizi, and M.E.J. Newman, Power-law distributions in empirical data. SIAM Review 51(4), 661-703 (2009).

Assignment

- Today, we decide students for the first sessions.
 - Session 1 and 2 are fixed.
- **All students should e-mail following information by 16 Oct.**
 - Student ID Number, Name, e-mail for slack
 - 3 candidate dates, and
 - 3 papers you want to present
 - Choose full papers (8 pages or more) in top conferences/journals published in recent 3 years (2012 – 2015)
 - Topics of the papers should be related to Web.
 - Include both basic and recent papers
- **Assignment & Schedule will be released on 19 Oct.**
- Note:
 - Students who select earlier day have more chance to present a preferred paper
 - Let me know, if you change the paper to present

Slack

- We use slack.com to announcement and chat in sessions
- I will check your questions and comments in sessions and mark them if they are “good” (Mark will become scores for your credits)
- After receiving your e-mail, I will invite you to webeng2015.slack.com
- Your name in slack should be “NAME-StudentID”