DD2476 Search Engines and Information Retrieval Systems

Course Analysis, ir14

Hedvig Kjellström and Johan Boye

This analysis has been performed by Hedvig Kjellström and Johan Boye and is based on an online student questionnare with 8 open questions, answered by 15 of the students, and on discussions with students during the course.

Course Data

The course was given for the fourth time; third time under its current code, during periods 3 and 4, 2014, with course leader Hedvig Kjellström. 94 students registered for the course, 79 completed some part of the course, and 59 are now finished. One student from last year also raised his grades by doing the lab assignments again this year. The teaching activities consisted of:

- 11 lectures of which three were given by Hedvig Kjellström, two by Johan Boye, two by Jussi Karlgren , one by Viggo Kann, and three by guest lecturers,
- Altogether 70 man-hours of computer hall sessions held by Hedvig Kjellström, Johan Boye, Carl Eriksson, Viktor Mattsson, and Gunnar Eriksson, where students presented the programming tasks.

The examination consisted of three computer assignments (6 hp) which were performed individually and examined orally in period 3, and one project (3 hp) which was performed in groups of four or five and examined by a interactive poster session and a written report in the end of period 4.

We are very happy to see that the course also this year was highly appreciated by the students. 14 out of the 15 students answering the questionnare were positive, with comments such as:
"*Awesome!*"
"*I liked the course design.*"
"*Overall I'm really satisfied with the course and impressed with the dedication of the teachers/professors!*"
"*Very interesting!*"

One out of the 15 student was more neutral/negative, and there were also general critizism against the dataset used for assignments (see below).

The book used in the course was (C. D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008). We also used (K. Avrachenkov, N. Litvak, D. Nemirovsky and N. Osipova, Monte Carlo Methods in PageRank Computation: When One Iteration is Sufficient, *SIAM Journal on Numerical*

Analysis 45(2), 2007), (S. E. Robertson and K. Spärck Jones, *Simple, Proven Approaches to Text Retrieval*, 1994), and (M. Sahlgren, *An Introduction to Random Indexing*, 2005).

Learning Outcomes

The course is based on Christopher D. Manning's book and course at Stanford. The first year, 2010/2011, we started off with a course that covered the whole book. In 2012 and 2013 we decided to shift the focus somewhat, from Machine Learning and XML (which is covered in other courses at CSC) towards Web Search, e.g., the PageRank algorithm (which is not covered by other courses). This year, we included a deeper focus on evaluation of IR systems, and picked in an expert in this area as a lecturer. The learning outcomes and course content were:

> After completing the course you will be able to:
>
> - explain the concepts of indexing, vocabulary, normalization and dictionary in Information Retrieval,
> - give an account of different text similarity measures, and select a similarity measure suitable for the problem at hand,
> - define a boolean model and a vector space model, and explain the differences between them,
> - implement a method for ranked retrieval of a very large number of documents with hyperlinks between them,
> - evaluate information retrieval algorithms, and give an account of the difficulties of evaluation,
> - give an account of the structure of a Web search engine.
>
> Content
> Basic and advanced techniques for information systems: information extraction; efficient text indexing; indexing of non-text data; Boolean and vector space retrieval models; evaluation and interface issues; structure of Web search engines.

Students were in general positive to the curriculum. On the question if they thought the curriculum was good or if there should be changes, they answered:
"*It was good.*"
"*I think, the course gave me a good introduction to search engines.*"
"*I thought it was great the way it is.*"

Some students student suggested more focus on Machine Learning aspects. This need is however covered by a new course in Machine Learning, DD2434, and will therefore not be added to the curriculum of this course:
"Topic modelling. What is the state of art in the IR." [Deep Learning for example, covered in DD2434]
"Is there any search engine that uses AI, pattern recognition or machine learning as a part of the search functionality? I would've been interested to learn more about that."
"*Some discussions about Machine Learning-based approaches and issues related to search engines would have been appreciated.*"

No action points

Course Activities

The course activities consisted of 11 lectures in period 3 and 4, and computer hall examination sessions in period 3 and beginning of 4. Lectures 1-7 covered the chapters in the same order as they appeared in the book, with the exception of Lecture 5 which reflected the strengthened web focus, and Lecture 3, which reflected the strengthened evaluation focus. Lectures 8-11 provided outlooks in related subjects such as Music and Image Retrieval, Evaluation, and IR design in practice:

Lecture 1: Given by Hedvig Kjellström and Johan Boye. Manning Ch 1, 2
Lecture 2: Given by Johan Boye. Manning Ch 2, 3
Lecture 3: Given by Jussi Karlgren. Manning Ch 8
Lecture 4: Given by Hedvig Kjellström. Manning Ch 6, 7
Lecture 5: Given by Johan Boye and Hedvig Kjellström. Manning Ch 21, Avrachenkov
Lecture 6: Given by Jussi Karlgren. Manning Ch 9, Robertson
Lecture 7: Given by Hedvig Kjellström. Manning Ch 11, 12
Lecture 8: Given by Viggo Kann. Sahlgren
Lecture 9: Guest lectures by Anders Friberg, KTH, and Hercules Dalianis, SU.
Lecture 10: Guest lectures by Simon Stenström, Findwise, and Magnus Rosell, FOI.
Lecture 11: Guest lecture by Filip Radlinski, Microsoft Research Cambridge, UK.

The guest lecturers from industry were a success this year as well, and will definitely be retained in the coming years. Some students raised complaints against the guest lectures. These will be taken into account when selecting future guest lecturers, but it is our firm belief that this is a highly industry-driven area of research and should not be taught as a purely academic subject:
"Hmmm.... The lectures were good, but most of the guest lectures were not well adapted for the audience."
"Please don't have the project being sponsored by a company. Keep this course academic and don't make it into some practical enterprise tutorial in big data."

No action points

Examination

As last year, the examination consisted of three computer assignments (6 hp, A-F) and one project (3 hp, A-F)

The assignments were designed to examine the part of the curriculum covered in Lectures 1-7. Grades were assigned based on how many tasks were performed (rather than how well the tasks were performed), which made made the grading transparent to the students. The grading also depended on if the assignment was presented on time; the grading system was explained clearly on the course homepage. The assignments were presented orally to a teacher in the computer hall. Students could book 15 minute time slots online for the presentation of each assignment.

As last year, the assigments were about building a rudimentary search engine. The assignments were arranged as:

Task 1.1: Basic Inverted Index
Task 1.2: Multiword Queries
Task 1.3: Phrase Queries

Task 1.4: What is a good search result?
Task 1.5: What is a good query? (C-A)
Task 1.4: Large Inverted Indexes (B-A)

Task 2.1: Ranked Retrieval
Task 2.2: Ranked Multiword Retrieval
Task 2.3: What is a good search result?
Task 2.4: Computing PageRank with Power Iteration
Task 2.5: No-Sinks PageRank Approximation (C-A)
Task 2.6: Monte-Carlo PageRank Approximation (B-A)
Task 2.7: Combine tf-idf and PageRank (A)

Task 3.1: Relevance Feedback
Task 3.2: Designing an evaluation
Task 3.3: Speeding Up the Search Engine (C-A)
Task 3.4: Ranked Bi-Gram Retrieval (B-A)
Task 3.5: Ranked Sub-Phrase Retrieval (A)

The students were generally happy with the computer assignments.
"*Overall, I liked the computer assignments very much..*"
"*They were really good!*"
"*Overall challenging and instructive*"

However there were quite serious problems with the dataset, due to different character encoding, sometimes many coding schemes within each document. This lead to problems in text matching. There were also issues relating to the language (Swedish – it becomes hard to evaluate if you cannot understand the document) and the link-structure (Wikipedia – very differently structured from the Internet overall). We will therefore exchange the dataset, and also look over the entire code skeleton to remove bugs:
"*The contents were often good, but there were many bugs in the instructions and the code skeletons."*
"*I'd like that the given code (thinking about the tokenizer) don't give different results depending on your OS configurations.*"
"*It would be nice if labs were in English :), but all in all they were ok and hard/easy enough.*"
"*In some cases I was not comfortable with the given program skeleton.*"
"*They were a little mazy, but after the first one I had a good understanding of what you were after. It really helped that both Hedvig and Johan were extremely helpful and quick to respond.*"

Based on the recommendations from last year's evaluation, a larger focus was put on evaluation; one or two evaluation tasks were added to each assignments. There were complaints, mainly due to the fact that the questions fitted badly to the dataset, which made the labeling and evaluation extremely unrewarding:
"*The tasks were good. The evaluation tasks were a bit annoying, since they took very much time and the search set was too small to find good information.*"
"*I really enjoyed all computer assignments, except for the parts where we had to do tedious manual query evaluation for a large number of query results... That part we could safely skip. I mean, you only need to do one such evaluation (of 1-5 documents, max) to grasp the concept. Other than that, the assignments were great.*"
"*I liked that there were elements of evaluation. However for some parts (e.g. the first assignment, Task 1.4 and 1.5), there was too much repetition. I at least skimmed through*

every article where I could not immediately be sure if it was irrrelevant from the title and that took a long time while I did not feel like I learned much more after doing this for two oder three queries."
"*I think the evaluation felt a bit pointless as the underlying dataset was not very good. Hardly any results were relevant.*"
While we will put some work into improving these exercises until next year (see, below), we will definitely retain the evaluation focus since this is an important part of the design – students have to learn this! Furthermore, many students found the evaluation exercises useful:
"*Very good. Probably the most useful part of the whole course actually.*"
"*I think it was boring, but useful. You need to have the evalutation mindset to develop great products..*"

In the questionnare, we posed the question if the students would prefer to work with a functioning engine such as Solr/Lucene, rather than implementing their own engine from the (quite rudimentary) code skeleton provided by us. As in previous years, there was an overall preference of keeping the current lab design where the students build their own engine:
"*I believe that building the index yourself if what the course should, mainly, be about. Sure, there should be some lab on using a fully functioning engine such as Solr/Lucene, but the most important is implementing the theory yourself.*"

The students were also required to perform a project (3 hp, graded A-F), examined by an interactive poster session (May 16) and a written report. Grading was done according to the following: Five compulsory and four additional criteria on the poster presentation and report were defined – for more information visit the course homepage. To pass, all compulsory criteria have to be fulfilled. Higher grades are reached by fulfillment of additional criteria.

The projects were defined by researchers from Findwise, Gavagai, and FOI. The project formulations were intentionally very brief, as specification of the project was part of the task. The students were divided according to grades on the computer assignments into 14 groups of 5-6 students, to make the groups as homogeneous as possible in terms of ambition level. The projects were distributed upon the groups based on student preferences and the requirement that projects should be well distributed on groups. The groups were supervised by the project proposers. Due to the large number of students we were again not able to find unique tasks to all.

We were very impressed with the results of many projects, as well as their poster presentations. As of May 16, the results of the project are: 24 A, 23 B, 21 C, 5 D.

The poster session was very successful this year as well – this presentation format will definitely be retained!

The students generally appreciated the project:
"*I liked the project a lot. I am not sure if choosing the project teams according to grades is the best way (if there is a best way). I had a good grade in the assignments, as did my teammates and we worked very well together. I feel like this might create some dissatisfaction for people that are completely able of completing all tasks in the assignments, but didn't have time due to stress in other courses.*"
"*It was fun.*"

"*Ok.*"
"*The group project was fun. I think it could have been better to start earlier. And maybe in smaller groups. With 6-7 members its easy for people to free-ride.*"
"*The project was also great, we enjoyed our project and it worked out perfectly*."
"*An overall good project that allowed us to dig deeper into different methods.*"
"I had a fun group, and we worked well together. It wasn't the most interesting assignment though, and it wasn't especially research oriented, which made it kind of disconnected from the rest of the course. However, I noticed that several other projects had a much stronger connection to the course material, so it definitely wasn't a problem overall."

We asked the students if they preferred the present very loose project definitions or if they would have liked to have more steering. There was a clear bias towards keeping the free format:
"*Yes, more structure and more definitions. Or just skip the project and make it into a lab exercise.*"
"*Perfect.*"
"*I liked that they were so loosely defined. We mostly still set out own deadline for steps in between, but it allowed us to adapt the schedule according to other responsibilities. This is less possible when there are many official deadlines in between.*"
"Subject was a bit loosely defined but the organization was our business so it did not really matter."
"I would not have liked project diaries and/or more reports! This was fine!"

The grade of the course is assigned as the weighted mean (rounded up) of the computer assignment and project grades. 59 students have finished all parts of the course, with the following grades: 22 A, 15 B, 11 C, 11 D. This is a bit to the high side, but we believe that it is not far from reflecting the true quality of the students, which were on average very good.

Action points:

We will rework the assignments, fixing bugs and improving the skeleton. We will also find and prepare a new dataset which 1) is in English, 2) is more similar to a general subset of the web, 3) with a contained link structure, almost isolated from the rest of the web.

We will improve the evaluation tasks in the computer assignments, using fewer test queries that fit better with the dataset.

We will think of a better method for finding project proposals, to enable every group to work on a unique problem. We failed to do this for this year, but it is a high-priority task for next year.