Assessing Interestingness and Importance of Information Retrieval Course Topics in a Course for Three Different Target Groups

Daniel Blank
Media Informatics Group
University of Bamberg
96045, Bamberg, Germany
daniel.blank@uni-bamberg.de

Andreas Henrich
Media Informatics Group
University of Bamberg
96045, Bamberg, Germany
andreas.henrich@uni-bamberg.de

ABSTRACT

Teaching an information retrieval (IR) course for three different target groups is challenging. These target groups are i) resident students at the University of Bamberg, Germany ii) remote full-time students from other Bavarian universities, and iii) remote but part-time students from all parts of Germany which usually study the course besides their day-to-day work relationship. As a consequence, we have participants with heterogeneous previous knowledge and potentially different expectations with respect to the course content.

In this paper we will only briefly describe the didactic aspects and challenges of the course. The clear focus of the paper lies on an in-depth quantitative evaluation how various IR topics presented in the lectures and tutorials fit the needs of the different target groups. We evaluate *interestingness* and *importance* of the course topics according to the students' impressions. The goal of the evaluation is to get hints for future refinements of IR course content—in general but also with respect to the needs of different target groups.

CCS Concepts

•**Information systems** → *Information retrieval;* •**Applied computing** → *Distance learning; E-learning;*

Keywords

Teaching Information Retrieval, Evaluation of Course Content, Distance Learning

1. INTRODUCTION

As a trend in recent years, part-time studies on the job and modular study programs become more and more important—also for universities. Often, when individual courses for different target groups are difficult to provide, existing courses must be tailored to the needs of heterogeneous target groups.

When it comes to teaching information retrieval (IR), questions arise how various target groups which study the course perceive different IR topics. This is interesting because of several reasons: First, it gives course providers a hint how to improve and extend their course in general. Second, different target groups bring differing previous knowledge with them and therefore heterogeneous views and expectations on the course topics. Obviously, target groups benefit from each other, for example when doing joint projects, working together when solving exercises and discussing topics. Therefore, it is a main challenge to attract all target groups and at the same time to deliver the basic IR topics in an adequate way.

This paper evaluates how various IR topics presented in lectures and tutorials fit the needs of three different target groups. We evaluate *interestingness* and *importance* of IR course topics according to the students' impressions. Our target groups are i) resident students at the University of Bamberg, Germany ii) remote full-time students from other Bavarian universities, and iii) remote but part-time students from all parts of Germany which usually study the course besides their day-to-day work relationship.

The paper is organized as follows. In section 2, we describe our course setting in more detail (target groups, course content, and the learning concept). Afterwards, in section 3, we present an evaluation how the different course topics attract the students. Finally, chapter 4 concludes this work.

2. CONTEXT AND COURSE CONCEPT

In this section, we first focus on the different target groups we address with our course. Afterwards, we present the learning concept on which our course is based and outline the course content.

2.1 Target Groups

The IR course at the University of Bamberg, Germany, is a masters course addressing three different target groups.

The first group of students are domestic students which study at the University of Bamberg (*ba*). Most of them take master courses at the Faculty of Information Systems and Applied Computer Sciences. In addition, there are students which come from other faculties of the University of Bamberg. All these students usually live in or close to Bamberg and thus have the opportunity to attend the lectures and tutorials offered in Bamberg on a weekly basis.

The second group of students comes from Bavarian universities other than the University of Bamberg. They are usually full-time students at their home university and study the course using the service provided by the Bavarian Virtual University (BVU)[1] (*bvu*). BVU students do not come to Bamberg to attend lectures and tutorials but use the support and e-learning facilities offered by the course providers. These facilities are of course also available for the domestic students.

The third target group are VAWi students (*vawi*)[2]. The VAWi degree is a Master of Science in information systems offered in cooperation by the University of Bamberg and the University of Duisburg-Essen. Usually, VAWi students are in a work relationship and thus study course content in the evenings, in holidays, and on the weekends. Thus, they also rely on the e-learning support from the course providers.

2.2 Learning Concept

The course is mainly designed as an e-learning course. Thus, we briefly describe the e-learning concept here. Not attending the lectures does not put those students who decide to do so in a worse position since the lecture recordings also cover questions of students and corresponding answers of the lecturer. The same holds for the weekly tutorials where no additional content is presented. Thus, the main purpose of the tutorials is to answer questions of the students which they have when working on the exercise sheets. Remote students can ask questions in the forum or per e-mail. This is also the case for domestic students who do not attend the tutorial and/or the lecture because of the e-learning character which does not force them to do so.

The e-learning concept is as follows. Lectures are recorded and can thus be viewed and downloaded as lecture recordings. In addition, tutorials are given with three exercise sheets per semester (each assuming 20 hours of work).

We provide the slides as PDF files, the recordings of the lectures taken with Camtasia[3] (both in Flash/HTML5 as well as MP4 format), supporting material and remarks to particular course contents, forums for the communication between teachers and students but also among students themselves. All these materials and communication is provided by two Moodle[4] instances, one for the VAWi students[5] and another one for the two remaining target groups[6].

A typical work model for the course, especially for the remote students, is as follows: slides are printed and together with the lecture recordings the students study the slides, make notes, ask questions in the forum, etc.

In figure 1 you can see a small screenshot of a lecture recording. It shows the Flash/HTML5 formats with possibilities for navigating through the slides (on the right) and a small video showing the lecturer (on the left). The MP4 files are additionally provided for the use on mobile devices.

2.3 Course Content

In the following two sections, we will briefly outline the content of the lectures and the exercises given in the tutorials.

[1] *https://www.vhb.org/en/homepage/*
[2] *http://www.vawi.de/*
[3] *https://www.techsmith.de/camtasia.html*
[4] *https://moodle.org/*
[5] *https://lms.vawi.de*
[6] *https://vc.uni-bamberg.de/*

Figure 1: Lecture recording (produced with TechSmith® CamtasiaStudio®)

2.3.1 *Lectures*

When designing the course, we analyzed IR topics which should be addressed by the lecture and by the tutorials [1, 4]. As a prerequisite we wanted to provide a course which is in large parts based on an introductory textbook. Our choice is [3] which is in our opinion good to assess for a broad audience, especially for students with no core computer science background. In addition, it is now freely available[7].

However, for our course we selectively pick some chapters from [3] and rearrange them together with additional chapters in an order we consider best suited for our students.

A syllabus of the course content is given in table 1. It also shows (in curly brackets the chapter numbers of [3]) that we address the same topics as proposed in the first eight chapters of [3], but in a different order. We also slightly modified the contents of these chapters.

The main reason for dealing with IR evaluation and IR models at an early stage is that we want to address these topics with the first two exercise sheets. As a consequence, other topics such as crawling are provided late in the course.

2.3.2 *Tutorials*

As mentioned before, there are three exercise sheets for the semester. The workload of each sheet is scheduled with approximately 20 hours per person. Besides the exercises described in the following, there are also plenty of small paper and pencil exercises for the students. These are usually very brief and should train the students with respect to the final exam, some examples are: calculations of TFIDF, BM25, etc. or evaluation measures such as recall & precision, NDCG, BPREF, etc. These paper and pencil exercises are however not evaluated here because this would blow up our small quantitative study. Instead, we only evaluate exercises of two different types: i) reading exercises and ii) programming exercises in Java[8]. Both exercise types can be found on the exercise sheets. In total we designed five of these exercises.

Fleiss (sheet 1) This small reading exercise addresses IR evaluation and inter-rater reliability. Students were asked to read the Wikipedia article on Fleiss' Kappa[9]. In addition, they were given some evaluation data and

[7] *http://ciir.cs.umass.edu/irbook/*
[8] *https://www.java.com/*
[9] *https://en.wikipedia.org/wiki/Fleiss'_kappa*

1 Search Engines and Information Retrieval {1}
2 Architecture of a Search Engine {2}
3 Evaluation of Search Engines {8}
- 3.1 Why evaluate?
- 3.2 Evaluation Corpora
- 3.3 Querylogs
- 3.4 Effectivity Measures
- 3.5 Efficiency Measures
- 3.6 Training, Testing and Statistics

4 Retrieval Models {7}
- 4.1 (Overview of) Retrieval Models
- 4.2 Probabilistic Models
- 4.3 Language Models
- 4.4 Complex Queries or "Combining Evidences"
- 4.5 Websearch
- 4.6 Machine Learning and IR

5 Ranking with Indexes {5}
- 5.1 Introduction on Indexes
- 5.2 A Ranking Model
- 5.3 Inverted Files
- 5.4 Compression
- 5.5 Auxiliary Data Structures
- 5.6 Indexing
- 5.7 Query Processing

6 Working with Text (Documents) {4}
- 6.1 From Words to Terms
- 6.2 Text Statistics
- 6.3 Parsing a Document
- 6.4 Document Structure and Markup
- 6.5 Link Analysis
- 6.6 Information Extraction
- 6.7 Foreign Languages – Internationalization

7 Queries / User Interfaces / Interaction {6}
- 7.1 Information Need
- 7.2 Search Situations and Search Tactics
- 7.3 Query Transformation and Processing
- 7.4 Result Presentation
- 7.5 Cross-lingual Search

8 Crawls and Feeds—or: What is when indexed? {3}
- 8.1 Web Crawler
- 8.2 Desktop Crawler
- 8.3 Document Feeds
- 8.4 Converting
- 8.5 Storing Documents
- 8.6 Duplicate Detection
- 8.7 Noise Removal

9 Web Intelligence, SEO, Enterprise Search, Expert Search {n/a}
- 9.1 Web Intelligence
- 9.2 Search Engine Optimization
- 9.3 Enterprise Search
- 9.4 Expert Search

10 Image Retrieval {n/a}

Table 1: Table of contents (TOC) of our course

had to calculate Fleiss' Kappa in order to judge interrater reliability.

ERR (sheet 1) This reading exercise followed the idea of reading a research paper with some impact on IR evaluation. We choose [2]. Students were asked to answer some questions on the paper (e.g. describing underlying retrieval models of different evaluation measures) and to calculate expected reciprocal rank (ERR) values for some given hypothetical search engine rankings.

Lucene (sheet 2) In this exercise students read the introductory chapter of Lucene in Action [5] which is freely available on the book website. Afterwards, students were asked to index the CACM collection[10] with Apache Lucene[11] and implement query processing using four different IR models that Lucene offers. In addition, students implemented a simple evaluation measure such as MAP in order to analyze the quality of different IR models w.r.t. the CACM collection. Hereby, also the effects of stemming were analyzed.

TopMod (sheet 3) The basis of this reading exercise was reading [6] to get a better understanding of probabilistic topic models which are only briefly covered by the lecture. After reading the paper, students were asked to use the latent Dirichlet allocation (LDA) library Mallet[12] to generate some term-topic and topic-document matrices for the small test document collection that comes with Mallet. In a second step, students had to calculate term-term, query-document, and document-document similarities according to the techniques described in the paper for a small toy example.

MapRed (sheet 3) In order to introduce functional programming concepts and the basic idea of MapReduce students were first asked to read the Java 8 tutorials on Streams[13]. Afterwards, a first introductory part of the exercise addressed a simple credit card revenue aggregation example as described in section 5.6.3 of [3]. As a second part, a text file of Shakespeare's Hamlet was used in order to preprocess the text, count term occurrences and finally sort them in descending order.

3. CONTENT EVALUATION

In this section, we focus on the evaluation. First of all, the survey design is addressed. Then, the preferences of the different target groups are analyzed.

3.1 Survey Design

For the evaluation we designed a questionnaire[14]. It consists of two types of questions. Free text questions are used to ask which course content was missing or should be extended. More focused questions are based on the course syllabus (see table 1) where for all chapters and sections of the lecture as well as for the exercises in the tutorials we asked the following two questions:

[10] *http://ir.dcs.gla.ac.uk/resources/test_collections/cacm/*
[11] *https://lucene.apache.org/*
[12] *http://mallet.cs.umass.edu/topics.php*
[13] *https://docs.oracle.com/javase/tutorial/collections/streams/*
[14] *https://www.soscisurvey.de*

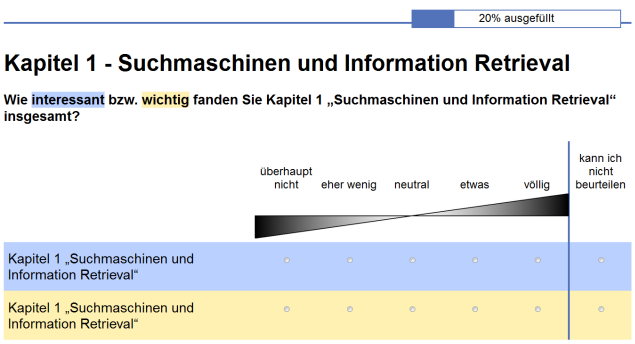

Figure 2: Questionnaire

How interesting ...?

Please state how interesting the course contents are according to your impression. Do the course contents inspire you and do you enjoy studying them?

How important ...?

Please state how important it is according to your impression that these course contents are part of the course. Importance addresses the attaining of a fundamental IR understanding but also the usefulness of the course contents in your future (working) life.

We decided to evaluate interestingness and importance in the way we did because of multiple reasons: In combination with the detailed table of contents (TOC) a more comprehensive questionnaire would possibly have lead to a too time-consuming completion with many aborts. Enjoying the studies and being inspired by the content is a key factor which we want to encourage. If content is perceived interesting, it is easier to deliver. Thus, the first question also tries to figure out if we deliver the content in the right way. In addition, importance should reflect if we deliver the right things. Of course, we know that this might be difficult to judge for the students. However, especially the VAWi students with a working background and students with interns might have some in-depth experiences in this regard.

Average completion time of the survey was 8.4 minutes with a median of 7.8 minutes. The questionnaire was accessible for two and a half weeks (10/2/16 - 28/2/16) close to the exam period so that students could on the one hand use the completion of the questionnaire as an exam preparation and on the other hand students were already deeply familiar with the course contents when answering the questions.

If a chapter in the TOC consists of multiple sections, the two questions of interestingness and importance were asked for the whole chapter but also for individual sections. Possible answers were *not at all* (1), *rather small* (2), *neutral* (3), *to some degree* (4), and *completely* (5). As a sixth answer opportunity students could indicate that they cannot judge the particular question (-1). A screenshot of the questionnaire—designed in German—is displayed in Fig. 2.

Thirty students took part in the evaluation: four (*vawi*), seven (*bvu*), and nineteen (*ba*) students[15]. Although we expected more participants, the numbers of participants per group are still enough to allow for a focused analysis of the results w.r.t. different target groups.

[15]For comparison, the number of participants in the final exam are eleven (*vawi*), nine (*bvu*) and forty-two (*ba*).

3.2 All Students

In a first analysis, we take a look at the overall outcome with results aggregated over all three target groups. For the lecture (Fig. 3, the number of data items contributing to a boxplot denoted in brackets, whiskers lead to the 1st/3rd quartile minus/plus 1.5 times the interquartile range, a plus indicates an outlier), it can be observed that on a global picture all aspects seem to be quite interesting and important with interquartile ranges lying between 3 and 5. In most of the cases, the median value is 4. According to interestingness, only sections 5.5 *Auxiliary Data Structures* and 9.4 *Expert Search* have a median smaller than 4 (*to some degree*). A closer look reveals that section 5.5 is a very short section with only a single slide on some basic data structures necessary to understand the general query processing. We cannot simply get rid of this. However, the section on expert search is more extensive and modular so that it might be replaced or restructured in future courses.

Another interesting view on the data gives us those sections which are both very interesting and very important (median value of at least 4.5 and interquartile range between 4 and 5 in both cases). These sections are: 4.5 *Websearch*, 4.6 *Machine Learning and IR*, 5.3 *Inverted Files*, 8 *Crawls and Feeds* (the only chapter in total with such a clear preference), 8.1 *Web Crawler*, 9.2 *Search Engine Optimization*. All these are core IR topics with a clear focus on the Web. As a consequence of the results, we believe that it might be necessary to present chapter 8 at an earlier stage in the course, also in order to be able to address this topic in the exercises (see answer to free text questions later on where this was requested).

Now we take a look at the exercises (Fig. 4, left). It can be clearly seen that the Lucene exercise is the most important and the most interesting exercise. The preferences of all other exercises indicate that it is worth considering them (median values of 4 and interquartile ranges between 3 and 4 or 3 and 5). We do not focus on the results in more detail here but later on when comparing the preferences of the different target groups.

3.3 University of Bamberg (ba)

If we address the domestic University of Bamberg students only (ba), we can see that this particular plot shows similar trends to the overall plot (Fig. 5 vs. Fig. 3). This is quite obvious since the 19 Bamberg students are by far the largest cohort. However, it is beneficial to take a closer look at their preferences. The interestingness of section 5.4 *Compression* is clearly worse than in the overall plot and compared to other sections of the lecture (interquartile range between 2 and 4.5). If we take the same criterion as before to analyze positive preferences (median value of at least 4.5 and interquartile range between 4 and 5 in both cases), especially chapter 1 as a broad introduction and chapter 2 with basic architectural issues seem to be very appealing. Remarkable is also the outstanding interestingness of 9.2 *Search Engine Optimization*. Also 6.5 *Link Analysis* has a very positive preference. On the other hand, the importance of 7.5 *Cross-lingual Search* is judged rather low being a candidate for exclusion. However, this section is to some degree interesting and also very short in total.

Preferences according to the exercises are given in Fig. 4 (second column). Of course, they are also similar to the

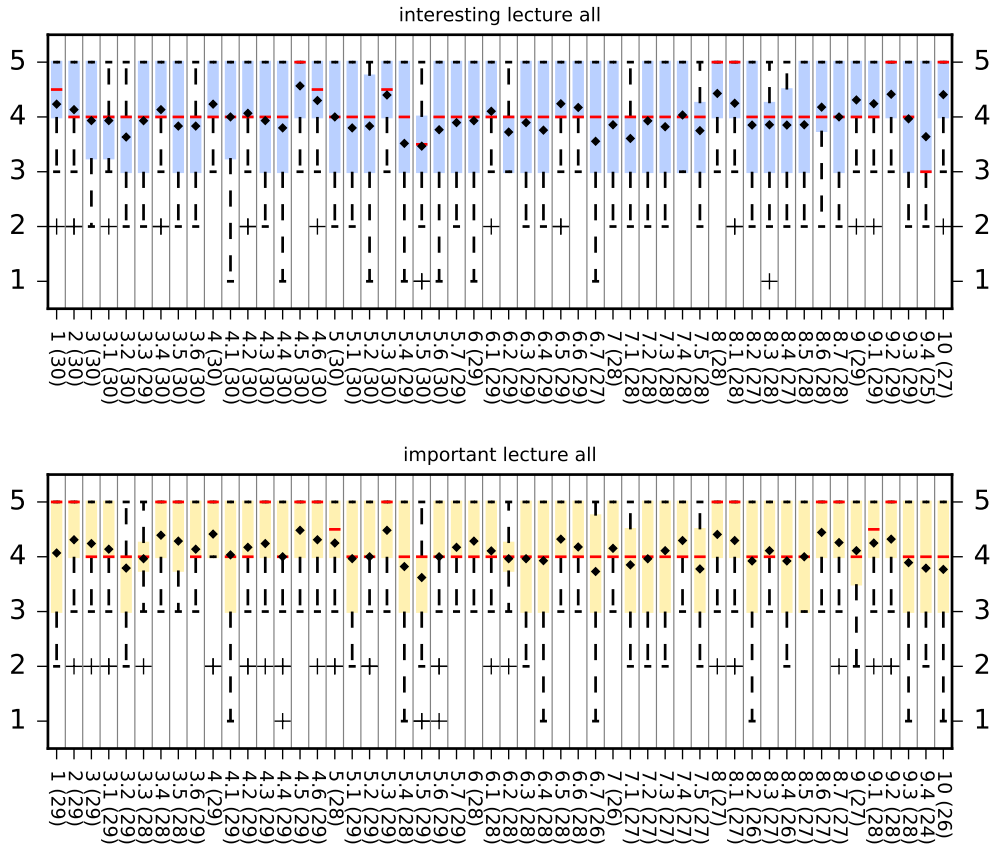

Figure 3: ALL: How interesting and important is the lecture?

overall trend with the Lucene exercise being the winner followed by the MapReduce task.

3.4 Bavarian Virtual University (bvu)

Let's take a look at the BVU cohort in Fig. 6. Here, the overall impression is worse with median values of 3 or 3.5 being more common. Students consider topics as being important and interesting, however only to some degree, especially with respect to the core IR topics addressed in chapter 4, 5 and 6. An exception from this trend is 3.6 *Training, Testing and Statistics*. It is questionable why this is the case since usually—for the other target groups—this section is of no superior interestingness. Students are usually familiar with these aspects because they are covered before in some statistics lectures. Maybe this is not the case for the BVU students, for which we do not have any insights into the study program. In total, it is somehow astonishing that more elaborate IR topics such as for example 4.3 *Language Models* and 4.4 *Complex Queries or "Combining Evidences"* do not attract these students to a larger extend.

For the exercises, it can be seen from Fig. 4 that the rather technical exercises with Java implementation aspects (Lucene and MapRed but also TopMod with a Mallet tutorial) are rated less important and interesting by the BVU group than by the others. These exercises are clearly rated as less interesting. Yet, their importance is perceived to some degree. Especially for the Lucene exercise this judgment is unexpected. Maybe it has to do with the previous knowledge in Java programming. Whereas for *ba* and *vawi* we know that the students are offered introductory Java courses and we know the course content, we do not know this for the BVU students, although it is clearly indicated in the course description that Java is compulsory. In total, similar to the lecture, the preferences of the BVU group according to the exercises are slightly worse than the global trend with results aggregated over all groups.

3.5 VAWi (vawi)

Finally, the *vawi* group—with a maximum of only four answers per question—rates the content of the lecture and the tutorials quite positive. Chapter 4 and in particular 4.2 *Probabilistic Models*, 4.3 *Language Models*, 4.5 *Websearch* and 4.6 *Machine Learning and IR* are rated highly interesting and important in Fig. 7. At least for 4.2 and 4.3, this is contrary to the preferences of the BVU group. In addition, chapter 5 (despite 5.6 *Auxiliary Data Structures*) and parts of chapter 8—in particular 8.1 *Web Crawler* and 8.6 *Duplicate Detection* with the best score—are judged very positive. Also chapter 10 *Image Retrieval* is rated highly interesting and important.

The judgment of 8.6 *Duplicate Detection* contrasts our impression. We would have considered this section being a topic which might be replaced in the future.

On the negative side, 7.1 *Information Need*, 7.2 *Search Situations and Search Tactics*, and 9.3 *Enterprise Search* can be named. A reason why 7.1 and 7.2 are rated so negative

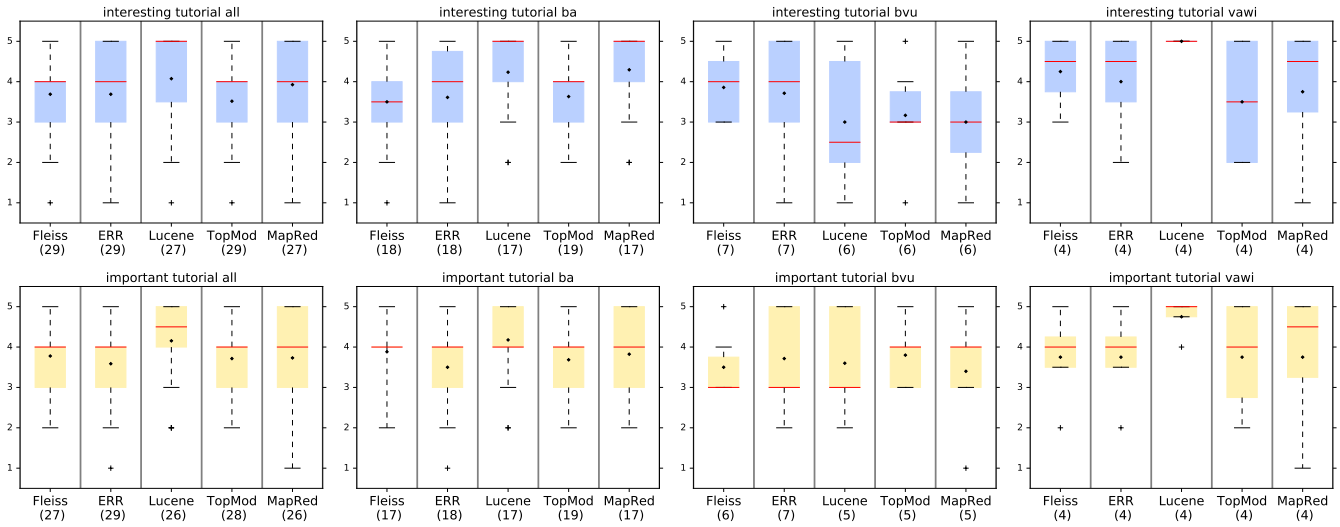

Figure 4: How interesting and important are the exercises of the tutorial?

might be that chapter 7 is a fundamental IR part typically presented at the beginning of a course or at least earlier than in our case. As we address it so late, several aspects have been briefly introduced or mentioned before (what is an information need, which search situations exist, etc.). Thus, we will have to prune course content here in chapter 7 in this regard or present the essential parts earlier in the course.

An interesting outcome of the tutorials evaluation—besides the overall positive impression—is that the Lucene exercise is seen so positive. Maybe this has to do with the working background of those students.

3.6 Free Text Questions

We also asked the students *which topics were missing* in the course. Three answers stated that nothing was missing at all. Reasons for that being that the course is already comprehensive and that there was not much previous knowledge. Since the answer to this question was not obligatory, possibly more students have this impression, without stating anything in the answer form.

As missing topics students stated:

- privacy and encryption and their influences on IR
- a particular section on implementation aspects with code snippets to the search algorithms
- Google architecture and Google and their algorithms should be assessed

In another free text question we asked *which content that is already addressed in the course should be presented more comprehensively.* Two students noted that there is no room for extensions. One particular student said that all aspects should be addressed more comprehensively. Some others gave the following answers (Note that we assigned some answers to the previous free text question to this question because they better fit in here. This is also a hint that we should clarify that these aspects are extensions of topics already present in the course.):

- Multimedia Retrieval (image, audio, and video) (5 mentions)
- Statistical Methods resp. Machine Learning in IR (2 mentions)
- SEO (2 mentions)
- Enterprise Search
- Web Search
- Crawling (should be addressed earlier and in exercises)
- NLP / how to deal with language in text
- Ontology-based Indexing

To address the first two aspects with 7 mentions we plan to extend our course in chapter 10 towards multimedia retrieval as an ideal playground for machine learning techniques such as for image classification and computer vision.

Addressing the tutorials and exercise sheets, we asked the following question: *From which topics would you like to have more exercises in the form of reading exercises and practical implementation exercises in Java?* Students gave the following answers:

- More programming exercises (2 mentions): e.g. extending the MapReduce exercise and weaving topics through the semesters; programming and comparing small ranking algorithms
- Apache Lucene (2 mentions): working on/with algorithms and experimentation
- Crawling (3 mentions): reading exercises and implementation
- More reading exercises addressing topics not covered in the lecture and current state-of-the-art research (3 mentions): e.g. evaluation, result presentation, query processing, NLP, machine learning.
- NER
- Web/SEO
- Google architecture
- I found the paper on ERR very interesting. It relates course content to actual work.
- Addressing further aspects is difficult because time is missing: The tutorial was however a good mix.
- I really liked the selection/mix of topics in the exercises
- The reading exercises were excellently shaped!

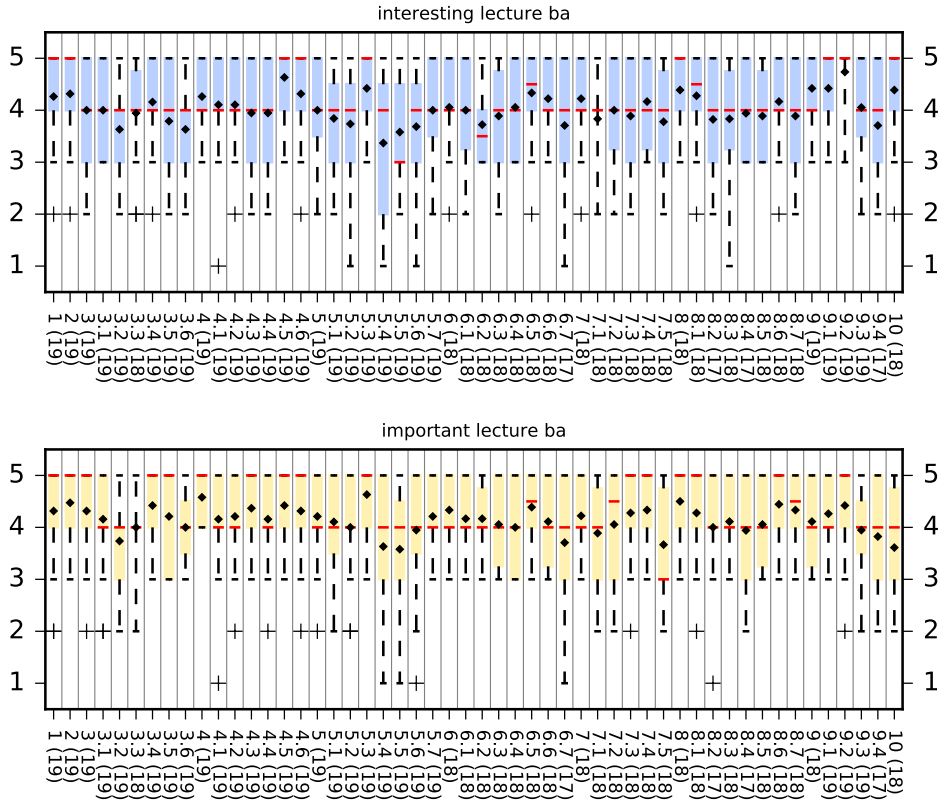

Figure 5: BA: How interesting and important is the lecture?

- To my impression, MapReduce was too much on Java. It was more focusing on stream capabilities in Java 8 than dealing with an IR system.

As a result, for the tutorial we plan even more programming tasks partially combined with reading exercises.

4. CONCLUSION

The small empirical study on the interestingness and importance of IR course contents reveals several interesting aspects. Although there lies more detail information in the data, some general aspects were shaped out here.

With respect to different target groups it is difficult to address all needs in a single, introductory IR course. Nevertheless, the study gave us some important insights on how to improve or restructure the course. First, crawling could be addressed earlier in the course and also in the exercises. Second, contrasting our thoughts on restructuring the course where we planned to shift away the strong focus on web search aspects, a comprehensive consideration of web search seems strongly desired.

For the exercises, the following can be concluded. Integrating reading exercises of state-of-the-art research papers fit the needs of the students and more of those kinds of exercises are requested. In addition, more programming exercises especially with a clear IR focus such as getting to know search engine libraries or crawling seems beneficial. In order to provide room for this, it might be a strategy to move the small exam preparation exercises which were not evaluated here to a special exam preparation exercise unit.

5. ACKNOWLEDGMENTS

We would like to thank all students who participated in the course evaluation and thus gave us helpful feedback on how to improve the course in the future.

6. REFERENCES

[1] D. Blank, N. Fuhr, A. Henrich, T. Mandl, T. Rölleke, H. Schütze, and B. Stein. Teaching IR: curricular considerations. In Efthimiadis et al. [4], pages 31–46.

[2] O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 621–630, New York, NY, USA, 2009. ACM.

[3] B. Croft, D. Metzler, and T. Strohman. *Search Engines: Information Retrieval in Practice*. Pearson, International Edition, Upper Saddle River, New Jersey, USA, 2010.

[4] E. N. Efthimiadis, J. M. Fernández-Luna, J. F. Huete, and A. MacFarlane, editors. *Teaching and Learning in Information Retrieval*, volume 31 of *The Information Retrieval Series*. Springer, 2011.

[5] M. McCandless, E. Hatcher, and O. Gospodnetic. *Lucene in Action*. Manning Publications, 2 edition, 2010.

[6] M. Steyvers and T. Griffiths. Probabilistic Topic Models. *Latent Semantic Analysis: A Road to Meaning.*, 427(7):424–440, 2007.

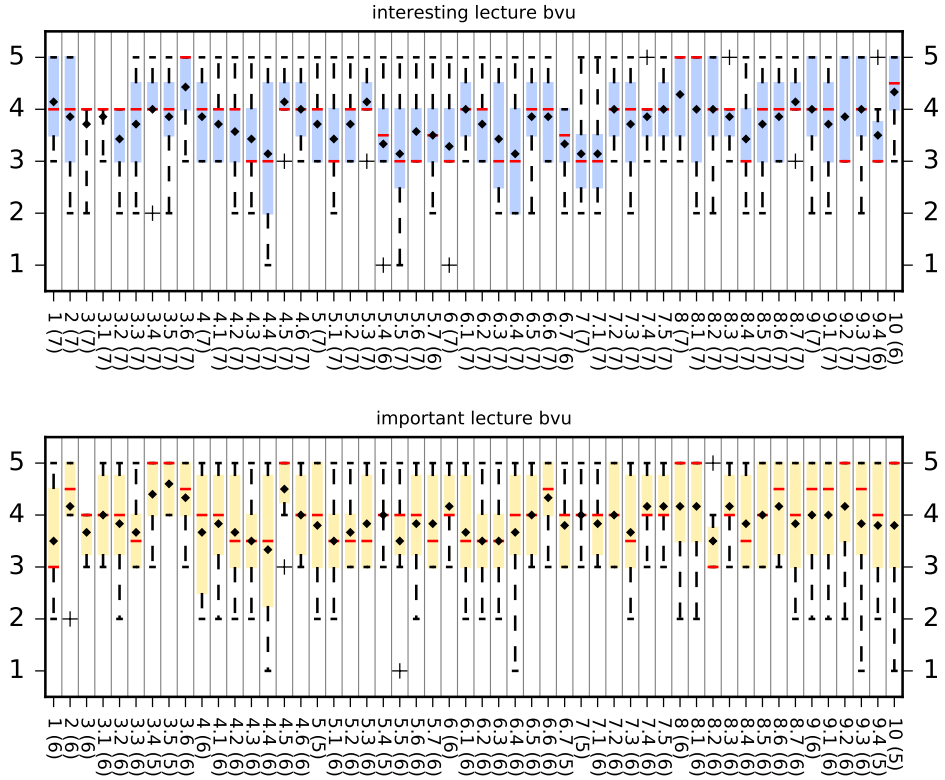

Figure 6: BVU: How interesting and important is the lecture?

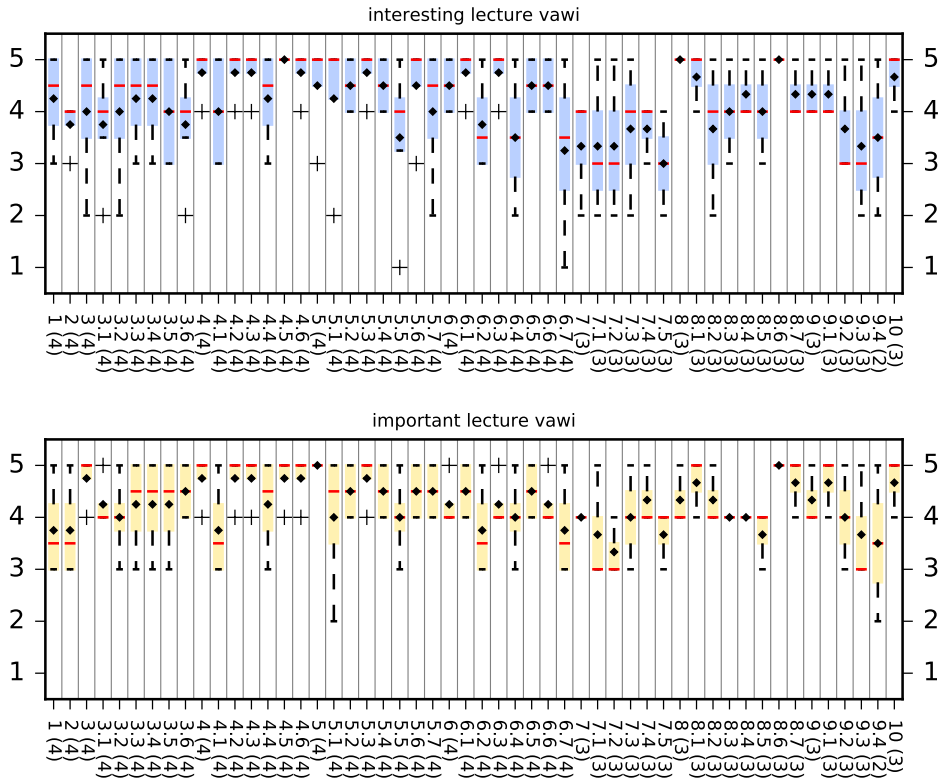

Figure 7: VAWI: How interesting and important is the lecture?