Information Retrieval

WS 2018 / **2019**

Lecture 14, Tuesday February 5th, 2019

(Course Evaluation, Exam, Work at our Chair)

Prof. Dr. Hannah Bast
Chair of Algorithms and Data Structures
Department of Computer Science
University of Freiburg

UNI FREIBURG

Overview of this lecture

UNI FREIBURG

- Organizational
  - Your experiences with ES13 — NLP, POS-Tagging
  - Official evaluation — results + discussion
  - Infos about the exam — when, where, how, tasks
  - Work at our chair — how, projects, next courses

Experiences with ES13

UNI FREIBURG

- Summary / excerpts
  - Not too many people did this last exercise sheet

    As usual, but it's nevertheless relevant for the exam
  - "The exercise sheet was a really fun and informative way to understand the Viterbi algorithm and how the core of NLP and NER work"
  - "It's a sad day today, because at this point I visited all of the lectures of the AD team and each lecture of these were the best lectures I heard by far"
  - Let's look at a demo of the **master solution** (from Claudius)

Results Course Evaluation 1/7

UNI FREIBURG

- Participants
  - Registered for exam: **51** … last year: 76
  - Participated in the evaluation: **46** … **great !**

    27 x Bachelor, 16 x Master, 3 x Other

    42 x Informatik, 4 x ESE, 0 x Other
  - Nominations for teaching award: **40** … **thanks a lot !**
  - In the following, a summary of your feedback
  - You find **all** the details linked on the course Wiki

Results Course Evaluation 2/7

UNI FREIBURG

- Style of the course … numbers from last year in grey
  - **Learned** a lot: 74% fully agree, 20% agree, 6% ok
    last year 67% 31% 2%
  - **Explained** well: 74% fully agree, 22% agree, 4% ok
    last year 73% 27% 0%
  - **Activity** asked for: 80% fully agree, 13% agree, 4% ok
    last year 76% 21% 4%
  - **Level** of contents: 58% appropriate, 42% high, 0% too high
    last year 48% 51% 1%
  - **Quality** overall: 76% very good, 22% good, 2% ok
    last year 73% 27% 1%
  - Criticism / suggestions: see slide 9

Results Course Evaluation 3/7

UNI FREIBURG

- Student's effort
  - Effort relative to ECTS … 1 = very high, 5 = very low

11% x 1	46% x 2	41% x 3	2% x 4	0% x 5	this year
22% x 1	41% x 2	36% x 3	0% x 4	1% x 5	last year
12% x 1	31% x 2	51% x 3	5% x 4	1% x 5	department average

  - Slightly more work than the average CS lecture

    In previous years, the number were more skewed towards high effort, but we have worked hard over the years to make the exercise sheets **both realistic, motivating and interesting, yet not too much work**

    (note that that is not easy at all)

Results Course Evaluation 4/7

UNI FREIBURG

- Materials and Online Support / Tutors
  - **Materials** helpful: 80% fully agree, 15% agree, 4% ok
    last year 72% 22% 5%
  - Consumed lecture by **presence or video recordings**:

    2% pres, 81% rec, 0% both, 16% other this year

    14% pres, 39% rec, 38% both, 10% other last year

    34% pres, 15% rec, 19% both, 32% other dep average
  - Video recordings: quality of the recordings much appreciated

    Mixed comments about the "recording only mode" this semester:

    About half of you said it didn't make a difference for them

    The other half said they would have liked to see the professor live and be able to ask questions, at least from time to time

Results Course Evaluation 5/7

UNI FREIBURG

- Assistant and Tutors
  - **Assistant** (Claudius Korzen)

    Claudius did a fantastic job running the course this year, based on the resources from last year (still a lot of work: forum, fixing errors, coordinating everything, …)

    Thank you very much!
  - **Tutors** (Patrick Brosi, Niklas Schnelle, Daniel Bindemann):

    Almost all of you were very happy with your tutors, in particular with their feedback, speed, friendliness, helpfulness

    Several of you explicitly thanked the whole team of the course

    A big thank you from my side to the team, too!

Results Course Evaluation 6/7

UNI FREIBURG

- Criticism / Suggestions
  - Most of the (not too many) negative comments were about the fact that there were **no live lectures** this time and no opportunity to ask questions live

    Note that you could have met your tutor any time
  - Besides that, there were the following suggestions:

    Deep learning methods for information retrieval

    A project at the end or as an extra course

    More info on what is actually used in popular search engines

    In some lectures, there was not enough time and some explanations had to be rushed in the end

Results Course Evaluation 7/7

UNI FREIBURG

- Planned improvements **from two years ago**
  - More theoretical sheets, in alternation with the (often more time-intensive) implementation sheet ✓
  - Consider the many notes taking after each lecture ✓
  - Clarify that test cases are mandatory, but also clarify that they do not have to be implemented 1-1 ✓
  - Clarify in advance which programming languages are appropriate for which exercise sheet ✓
  - Finish Python cheat sheet and make it available early ✓
  - Offer live tutorials again, until nobody comes anymore (2 years ago: one live tutorial, with meagre attendance) ✗

Exam 1/7

the new IDs don't have a photo anymore (makes sense)

UNI FREIBURG

- Where, when, what to bring
  - Oral exams (Freiburg B.Sc. Computer Science students only):

    February 13+19+25+26 and March 6, 30min in 51-02-28
  - Written exam (all other participants):

    February 12 (Tuesday), 14:00 – 16:30 in 101-026 + 101-036
  - Please bring: student id, colored pens, brain

    student id: make sure you look like your photo or vice versa

    colors: greatly improves readability for examples / drawings

    brain: greatly improves quality of answers in general

Exam 2/7

UNI FREIBURG

- Type of exam
  - The written exam is **open book**

    That is, you can bring books, paper, etc ... but please be ecological when printing out slides + good for your karma

    Electronic devices of any kind are obviously not allowed
  - In the oral exam, ask us if you are missing some detail

    But there is no time to start understanding things for the first time, during neither the oral nor the written exam
  - There will be a sub-forum for questions about the exam

    Answer speed will be slightly slower in the semester break

    Don't ask all your questions in the night before the exam

Exam 3/7

UNI FREIBURG

- Types of questions
  - **Type 1:** Do algorithm (or variant thereof) by example, like we often did in the lecture see **colored pens**
  - **Type 2:** Implement a small function (in Python, Java or C++) small indeed, at most 10 lines per task
  - **Type 3:** Small calculations or proofs see **brain**
  - In general: the emphasis is on **understanding**, not on learning things by heart
  - Important: the contents + insights from the **exercises** are (very) relevant for the exam

Exam 4/7

UNI FREIBURG

- Preparation
  - To check whether you understood something:

    Put away the material from the lecture and try to write it down / prove it **in your own words / formalism**

    There is no point in learning the individual steps of a proof or argument by heart … it doesn't work that way
  - Once the basic stuff from the lecture is understood, the best preparation is to do all the exercise sheets

    If you work your way through all the exercise sheets (yourself), there is no way you can fail the exam
  - Also: work through the old exams (linked on the Wiki)

Exam 5/7

- Task 1.1 from WS 15/16 exam
  - Order the given documents according to the given query, and compute the AP for the given relevances

	D_1	D_2	D_3	D_4	D_5	Q
bla	1	1	2	2	3	1
blu	1	0	1	2	2	0
blu	0	1	1	0	1	1
	1	2	3	2	4	

Dot-Product scores

Ranked: REL D_5, D_3, D_2, D_4, REL D_1

swap

1 0 0 0 1

$R = 2$ (#relevant docs)

$R_1 = 1$ (position of first rel. doc)

$R_2 = 5$ (position of sec. rel. doc)

$P@R_1 = P@1 = 100\%$

$P@R_2 = P@5 = 2/5 = 40\%$

$AP = \frac{100\% + 40\%}{2} = 70\%$

Exam 6/7

- Task 1.2 from WS 15/16 exam
  - Write a function that computes the average precision given a bit vector as described for Task 1.1

bit array from 1.1
e.g. 1 0 0 0 1

```
def compute_ap (rels):
    prec_sum = 0          # sum of the P@R_i
    rel_num = 0           # number of rel docs
    for i, rel in enumerate(rels):
        if rel == 1:
            rel_num += 1
            prec = rel_num / (i+1)
            pre_sum += prec
    return prec_sum / rel_num
```

Exam 7/7

G

UNI FREIBURG

- Task 1.3 from WS 15/16 exam
  - Prove that AP = 100% if and only if all the relevant docs are ranked before all the non-relevant documents

r = # relevant docs

"$\Rightarrow$" $AP = 100\% \Rightarrow P@R_r = 100\%$

$\Rightarrow$ The first r docs are all relevants

"$\Leftarrow$" The first r docs are all relevants ("1...1 0...0", R_1 … R_r, 1 … r)

$\Rightarrow R_1 = 1, R_2 = 2, \dots, R_r = r$

$\Rightarrow P@R_1 = 100\%, P@R_2 = 100\%, \dots, P@R_r = 100\%$

$\Rightarrow AP = 100\%$

Work at our Chair 1/5

and we have many really interesting and large **datasets**

UNI FREIBURG

- How we work
  - We solve practically relevant (usually hard) problems

    Route Planning, Transit Maps, Search As You Type, Semantic Search, Question Answering, PDF extraction, …
  - We use theory as a (vital / essential) tool

    All our work has a solid theoretical understanding, otherwise solving complex problems remains hacking and guesswork
  - We make our software + results available to the public

    This requires an effort to write good software, good documentation, nice user interfaces, and so on …

    All three aspects were clearly visible in this course, too !

Work at our Chair 2/5

UNI FREIBURG

- Supervision
  - Similarly as in the lecture:

    Very good infrastructure + support, but apart from that you can (and are supposed to) work very independently

    Team work is, of course, highly desirable and encouraged

    Great for enthusiastic people who care about practical stuff and who want to get things done

Work at our Chair 3/5

UNI FREIBURG

- Machine Learning
  - We are building more and more on **machine learning** to solve our problems

    Not because it's fashionable … but because it's practical

    It's quite obvious that learning is the future for problems like natural language understanding
  - You have seen a few learning algorithms in this lecture:

    k-means, Naive Bayes, LSI, Hidden Markov Models, …

    For a few years now, we also use **deep learning**

    The complexity lies not so much in the algorithms, but in understanding how and why they work how well

Work at our Chair 4/5

UNI FREIBURG

- Current projects and demos
  - Route planning (part of Google Maps) demo
  - World-wide public transit visualization (Travic) demo
  - Finding the correct vehicle paths (Pfaedle) demo
  - Automatic drawing of nice transit maps (Loom) demo
  - Large-scale SPARQL+Text search (QLever) demo
  - Search-as-you-type semantic search (Broccoli) demo
  - Question Answering (Aqqu) demo
  - Text extraction from PDF (Icecite) paper
  - Super-easy availability via **Docker** Wiki

Work at our Chair 5/5

UNI FREIBURG

- Upcoming courses
  - **Algorithms and Data Structures** … in the SS 2019

    Basic course for B.Sc. Informatik+ESE students
  - **Information Retrieval** … in the WS 2019/2020

    You know it … become a tutor if you do a great exam !
  - **Programming in C++** … in the SS 2020

    The programming course for the whole faculty
  - **B.Sc. / M.Sc. projects or theses**

    Offered all the time, read the instructions on our **Wiki** and then just ask (if no response, ask again please)

    It helps if you attended one my lectures, with a good grade