Exam DM840 Cheminformatics (2014)

Time and Place

Time: Thursday, January 22, 2015, starting XX:00.
Place: The exam takes place in U-XXX

Even though the expected total examination time per student is about 27 minutes (see below), it is not possible to calculate the exact examination time from the placement on the list, since students earlier on the list may not show up. Thus, students are expected to show up plenty early. In principle, all students who are taking the exam on a particular date are supposed to show up when the examination starts, i.e., at the time the first student is scheduled. This is partly because of the way external examiners are paid, which is by the number of students who show up for examination. For this particular exam, we do not expect many no-shows, so showing up one hour before the estimated time of the exam should be safe.

Procedure

The exam is in English. When it is your turn for examination, you will draw a question. Note that you have no preparation time. The list of questions can be found below. Then the actual exam takes place. The whole exam (without the censor and the examiner agreeing on a grade) lasts approximately 25-30 minutes. You should start by presenting material related to the question you drew. Aim for a reasonable high pace and focus on the most interesting material related to the question. You are not supposed to use note material, textbooks, transparencies, computer, etc.

You are allowed to bring keywords for each question, such that you can remember what you want to present during your presentation. As a guideline you are expected to not have more than 10 keywords per question on the list, that you bring to the oral exam. Note that this list is expected to be put on the table during your oral exam.

We, the examinator and the censor, will supplement with specific questions when appropriate, and after a while, we will end the discussion of the exam question that you drew and turn to material from other parts of the curriculum. Note that all of this as well as discussion between examinator and censor about the grade is included in the approx minutes, so do not count on more than 10 minutes for your own presentation.

Some of the questions below are quite broad, so you must select the material you choose to cover. You will of course also be evaluated based on your selection of material. If you only present the simplest material, you limit the grade you can obtain. On the other hand, a good presentation of the simple material is better than a very poor presentation of the harder material. We might of course still ask you questions about material that you have decided to skip.

Curriculum

The curriculum consists of all documents marked as "mandatory" in the Blackboard System. This includes the following articles, book chapters, and slide sets (details on parts to focus on and parts that can be left out are given in the Weekly Notes).

```
Slides / notes:
---------------
week37-GraphGrammars.mandatory.pdf
week37.Isomorphism-Ullmann.mandatory.pdf
week38-mckay-notes.mandatory.pdf
week39.traces.partially-mandatory.pdf
week40-ring.partially-mandatory.pdf
week41.cycle-basis.mandatory.pdf
week43-petri-final.mandatory.pdf
week46-mol-des.partially-mandatory.pdf
week47-48.sysbiologiy.mandatory.pdf
week47-SMILES-notes.mandatory.(or weininger article).pdf
week50.polya-counting.mandatory.pdf


Book Chapters:
--------------
1124-Pallson-chapters_6_7_8_9_15-partially-mandatory.pdf
1127-AnalysisChapter3GF.mandatory.pdf
1210-AnalysisChapter5AnalyticCombinatorics.mandatory.pdf


Original Articles:
------------------
0908-ullmann-mandatory.pdf
0911.FlowILP.mandatory.pdf
0930-ring-perception-mandatory.pdf
1110-pca-mandatory-page_1-6.pdf
1116-weininger.87.smiles1-mandatory(or notes).pdf
1116-weininger.87.smiles2-mandatory(or notes).pdf
1217-cardelli-Artificial-Biochemistry-2009.partially-mandatory.pdf
```

In addition the curriculum contains the text of the required assignments and the Weekly Notes.

Exam Questions

In the following the list questions that you draw from is given (in bold face), (the list of subquestions is incomplete and just a suggestion.)

- **Canonical Representations**
  - Morgan's algorithm
  - SMILES notation
  - Canonical labeling of graphs (McKays algorithm (nauty), Traces)

  What is the computational complexity of the Graph Isomorphism Problem?
  What is the computational complexity of the Subgraph Isomorphism Problem?
  Explain Morgan's algorithm.
  What is a automorphism group? How is it used in traces/nauty?
  How do traces and nauty differ?

- **Generative Chemistries**
  - Graph Grammars
  - Double pushout approach
  - ILP approach in hypergraphs
  - Autocatalysis

  When expanding a chemical space with the DPO, what kind of isomorphism problems have to be solved, how many of them?
  Express autocatalysis as a "flow" problem in a hypergraph.

- **Ring Perception**
  - Hanser Algorithm
  - Cycle Bases in chemistry
  - Kirchhoff-fundamental cycle basis
  - Minimal cycle basis
  - Horton's algorithm
  - de Pina's algorithm

  What is a fundamental cycle wrt a spanning tree?
  What is a relevant cylce?
  Which ring/cylce sets are unique?
  Why is uniqueness important?
  How many relevant cycles can there be in a graph?
  What is the runtime of Horton's algoritm and de Pina's algorithm as discussed in the lecture?

- **Graph Isomorphism**
  - Subgraph and graph isomorphism (Ullmanns algorithm)
  - Graph isomorphism (McKay's algorithm (nauty), traces)
  - Relation to (generative) chemistry

Can Ullmann's algorithm be used to solve the subgraph- or the graph-isomorphism problem?

- **Petri Nets**
  - Properties (liveness/boundedness/reversibility)
  - Reachability graph, coverability graph
  - Invariants and how to determine them (Farkas algorithm)
  - NPc proof of reachable marking
  - Petri nets to model chemical/biological networks

  What is the output of Farkas algorithm? What is the runtime?
  When modeling Petri Nets with incidence matrices, is there information loss?
  Explain $M_0 + C \cdot u$.

- **Molecular Descriptors and QSAR**
  - Molecular descriptors
  - QSAR
  - Principal component analysis (PCA) / PCR

  How is P chosen in order to change the basis of data?

- **Mathematical Concepts in Systems Biology**
  - Stoichiometric models
  - Ths subspaces of S
  - Flux balance analysis
  - Flux cone / extreme pathways

  What is an extreme pathway?
  What are the differences between a convex and a linear space?

- **Combinatorics and Polya Counting**
  - Generating Functions
  - Symbolic Method
  - Cylce Index Polynomial and Burnside's Lemma
  - Connection to Chemistry (counting isomers, counting chlorinated compounds)

  How many chlorinated versions of Benzene are there?
  Having 2 combinatorial classes A and B, when one looks at the combionatorial class $A \times B$, what happens to the generating functions? Proof!
  Exercises from the exercise sheet!

- **Process Algebras**
  - π-Calculus
  - Modeling of Chemistry with stochastic π-calculus
  - Modeling of epidemic systems with stochastic π-calculus