

RUBRIKER FÖR KANDIDATAVHANDLINGAR I DATAVETENSKAP

18.1.2011

ALGORITMER, BERÄKNINGSBARHET OCH BERÄKNINGSKOMPLEXITET

P=NP PROBLEMET: TILLVÄGAGÅNGSSÄTT, MÖJLIGA SVAR, FÖLJDER

(THE P=NP PROBLEM: APPROACHES, POSSIBLE ANSWERS, CONSEQUENCES)

The P=NP question is perhaps the greatest open problem in Computer Science, that has intrigued researchers for decades. Despite abundant research in computational complexity and it being placed on the 1 million dollar problem list of the Clay Mathematics Institute ([http://www.claymath.org/millennium/P vs NP/](http://www.claymath.org/millennium/P_vs_NP/)), it has not yet been solved.

The student taking up this topic would give an introduction to the formulation of the P=NP problem and explain its relevance for theoretical and applied computer science. The thesis would then present the ideas of some of the (failed) approaches that have been proposed for this problem. Finally, the thesis would discuss the future of the problem in terms of possible answers and their consequences. When discussing the topic, students may choose their own favorite NP-complete problems such as SAT, shortest common superstring, or, why not, the Minesweeper game (see R.Kaye, 'Minesweeper is NP-complete', Mathematical Intelligencer 22 (4), 2000, pages 9-15).

Handledare: **Ion Petre**

DATABEHANDLING MED BIOMOLEKYLER

(COMPUTING WITH BIOMOLECULES)

Proposed less than a decade and a half ago, this is a new, elegant, and promising paradigm of computing using biomolecules (DNA, RNA, proteins) rather than electronic computers. Many experiments have been realized, including solving in a bio-lab some difficult (NP-complete) computational problems, or implementing games (such as poker or tic-tac-toe) using DNA. Many experimental techniques currently in use for nano-level programmable self-assembly are based on biomolecules designed in such a way as to assemble in the desired pattern.

This project would introduce the concept of how to compute with biomolecules and compare its computational power with respect to that of electronic computers. The student would review some of the recent experiments, their techniques, and the computational problems solved in this way. This will also include the computer-simulation of such experiments. Based on these simulations, one may investigate the perspectives of the current technologies in DNA-computing, including speed, scalability, and costs.

Handledare: **Ion Petre**

KVANTUMALGORITMER

(QUANTUM ALGORITHMS)

Quantum physics is an elegant theory that describes Nature at the level of elementary particles. In the 1980s a computing paradigm based on the principles of quantum physics was proposed and the concept of quantum computing was born. It turned out quickly that quantum computing is radically different than computing based on classical physics. In particular, they can be exponentially more powerful. Consequently, in a world with quantum computers, the current cryptographic basis for banking or for electronic commerce would have to be revisited.

The student taking up this project would discuss in his/her thesis the concept of quantum computing, including the very basic features of quantum physics necessary for understanding it. It would then review some of the most well known quantum algorithms such as that for factoring integers, or protocols for quantum cryptography. The thesis would also discuss the current state of the art in building quantum computers and perspectives for a world with quantum computers.

Handledare: **Ion Petre**

BERÄKNINGSBARA HEURISTIKER FÖR NUMERISK MODELLANPASSNING

(COMPUTATIONAL HEURISTICS FOR NUMERICAL MODEL FITTING)

A mathematical model associated to a biological (or physical, or chemical, or economical, etc.) model assigns a time-dependent variable to every actor of interest and describes the interactions between the actors in terms of mathematical relations. Very often, the mathematical relations have the form of systems of differential equations, but they can also be Markov chains, discrete dynamical systems, Petri nets, etc. The intensity of each interaction is controlled through a kinetic constant, whose value is often unknown. Estimating the values of these constants so that the numerical behavior of the model confirms a given set of experimental data is called numerical model fitting.

The student taking up this project would first discuss some of the principles of constructing a mathematical model associated to a given biological model. (S)he would then discuss some of the heuristics methods for numerical model fitting, such as particle swarm, simulated annealing, genetic algorithms, evolutionary algorithms, random walk, gradient-based methods, etc. Some of these methods could then be implemented and their performance on a benchmark model should be compared. The thesis could also discuss the parallelization of these algorithms.

Handledare: **Ion Petre**

BERÄKNINGSMETODER INOM SYSTEMBIOLOGI

EN GRAFBASERAD SIMULATOR FÖR GENSAMMANSLAGNING

(A GRAPH-BASED SIMULATOR FOR GENE-ASSEMBLY)

Gene assembly in ciliates is one of the most elegant examples of "computations" taking place in living organisms. Ciliates (some 2.5 billion years old unicellular organisms, including the fastest living form on Earth) apparently implement some involved pattern-matching techniques in a stage of their reproduction process. In fact, this is often considered to be the most involved DNA manipulation

process known in Nature! We focus in this project on a computational model for gene assembly consisting of only three rewriting rules describing the transformation of a micronuclear gene into a macronuclear gene. Extensive expertise on this model exists at Åbo Akademi, see http://combio.abo.fi/projects/gene_assembly.php.

This project would extend the currently existing Gene Assembly Simulator (<http://combio.abo.fi/simulator/simulator.php>). It would incorporate into the simulator the option to draw a directed graph and it would implement the simple gene assembly operations on directed graphs. The student taking up this project would also design and implement graph algorithms for analyzing the computational power of the simple gene assembly operations.

Handledare: **Ion Petre**

BIOMODELLERING MED PETRINÄT

(BIOMODELING WITH PETRI NETS)

Petri nets have been proposed for reasoning about distributed systems. There is a great amount of techniques, theory, and practical examples for how to model and reason with Petri nets. More recently, Petri nets have also been adapted to model biological networks and many of the Petri net tools developed in Computer Science for concurrent systems have been adapted for biology.

This thesis would introduce the basic notion of Petri nets and some of the most widely used of its variants. Some of the basic methods for Petri nets such as execution semantics, invariants, reachability, liveness, boundedness, etc. It would then discuss how biological models can be modeled with Petri nets and what kind of biological properties can be captured through analyzing the Petri net model. The thesis would also compare Petri nets with other methods of biomodeling such as differential equations or Markov chains. Finally, the thesis would consider a biological case study, such as the heat shock response, and analyze it through a Petri net model.

Handledare: **Ion Petre**

REGELBASERAD MODELLERING FÖR SJÄLVASSEMBLERANDE PROTEIN

(RULE-BASED MODELING FOR PROTEIN SELF-ASSEMBLY)

Self-assembly is a process through which some elementary blocks assembly into larger and larger complexes depending on local rules describing matching configurations. Examples of self-assembly processes are abundant in molecular biology. We focus in this project on protein self-assembly, with a particular stress on vimentin proteins. They self-assemble in the living cell to yield “rope-like” intermediate filaments.

A major difficulty in modeling protein self-assembly is handling the size of the model. Since protein complexes can in principle become arbitrarily large, the model has to be able to describe the indefinite enlargement of such complexes, thus leading to models of infinite size in terms of variables and interactions. To overcome this difficulty, the self-assembly is described through presenting the local configurations making the ends of two filaments to fit together. In computational terms this can be described through local rewriting rules.

The student taking up this project would introduce the principle of rule-based modeling in biology, with a special focus on the software environment at <http://www.rulebase.org/> (Harvard University).

The thesis would then construct, simulate, and analyze a rule-based model for intermediate filament assembly. Finally, the student would compare the performances of his/her rule-based model with those of an existing differential equation-based model.

Handledare: **Ion Petre**

BERÄKNINGSMETODER FÖR PROTEINVIKNING

(COMPUTATIONAL METHODS FOR PROTEIN FOLDING)

This problem is often called the "Holy Grail of Bioinformatics" and is currently one of the most actively studied topics (including a yearly international contest on the best prediction programs). In short, the problem is to predict the three dimensional folding of a given protein. Although our understanding of this folding is still limited, many models exist for this problem, combining various hypothesis and insights into the biochemistry of protein sequences with elegant algorithms on strings.

This project would consider some of the existing computational models for protein folding and implement them efficiently. Ideally, a visual output should be provided.

Handledare: **Ion Petre**

DIGITALA MEDIER

DIGITAL PUBLICERING I FORM AV E-BÖCKER

Den studerande skall ge en kort historisk överblick över e-bokens utveckling, men därefter fokusera på att presentera de e-bokstandarder och -läsare som är vanligast i dag och beskriva skillnaderna mellan dessa. Avhandlingen skall också adressera frågor kring immaterialrätt, licensering och kopieringsskydd som aktualiseras i samband med e-böcker.

Handledare: **Linda Mannila**

E-UNDERSKRIFT

Handledare: **Ulla Solin**

DISTRIBUERADE SYSTEM, NÄTVERK

AD-HOC NÄTVERK

Uppsatsen kunde t.ex. behandla ad-hoc nätverk i hemmamiljö.

Handledare: **Kaisa Sere**

KOMMUNIKATIONSMODELLER

(COMMUNICATION MODELS)

Communication takes place nowadays at a multitude of levels, among people, among devices, or among people and devices. Communication can be synchronous or asynchronous, technologically simple or sophisticated.

In a BSc/MSc thesis one can survey communication models, analyzing them for a set of properties, comparing them or case studying them. A thesis involves first the description (together with the afferent study) of the communication model, then the explanation of the investigated problem in the thesis, then the actual analysis/survey/comparison.

Handledare: **Luigia Petre**

WLAN-ROAMING

Handledare: **Ulla Solin**

GRID COMPUTING

En grid är ett distribuerat system som utnyttjar geografiskt distribuerade resurser för att utföra beräkningar beaktande resursernas tillgängighet, kapacitet, prestanda mm. I uppsatsen kan man fokusera på någon intressant aspekt inom grid computing, t.ex. skedulering och administration av resurser eller olika tillämpningsområden för grid computing.

Handledare: **Marina Waldén**

SJÄLVHELANDE SYSTEM I NÄTVERK

Självhelande system är ett system som kan upptäcka att det inte fungerar korrekt och utan mänskligt ingripande kan utföra nödvändiga justeringar för att återgå till fungera normalt. I uppsatsen kunde man t.ex. koncentrera sig på självhelande system som stöder nätverkskonfigurationer.

Handledare: **Marina Waldén**

KOLLABORERING I ANSEENDE-BASERADE SYSTEM

Hur och på vilka premisser kollaborerar aktörer över nätverk och hur förändrar den anonymitet nätverket medför en aktörs handlingar. Exempel på sådana system är Ebay (reputation score), Slashdot (karma), P2P-nätverk osv.

Handledare: **Mats Neovius**

MODELLERING AV KONTEXT-MEDVETNA APPLIKATIONER

Hur kan man modellera kontext-medvetenhet (context-awareness) samt vilken är dess roll i applikationer. Varje applikation exekverar för något ändamål och kräver därmed någon input från den informella världen, uppsatsen skulle söka svar till hur dessa kontext beronden kan (har) modellerats

Handledare: **Mats Neovius**

TILLIT PÅ KONTEXT UR DATORMOLN

Datormoln (cloud computing) baserar sig på Software as a Service. Varje service kan antas ett inkapslat fragment av ett program, hur kan tillit uttryckas och vilka är svagheter för detta. Hur uttrycks tillit på kedjor av kontext.

Handledare: **Mats Neovius**

FORMELLA METODER

FRÅN AKTIONSSYSTEM TILL EVENT-B

(FROM ACTION SYSTEMS TO EVENT-B)

Action Systems and Event-B are two languages for precisely specifying software and systems. Action Systems is more general and flexible than Event-B, while Event-B has an associated proving tool – the RODIN platform.

In a BSc thesis one can compare these languages, from certain points of view. Additionally, in a MSc thesis, one can investigate manual and automated translations between systems specified using either language. A thesis involves (studying and) describing the two languages, then explaining which features will be compared (and potential reasons for that), then the actual comparison.

Handledare: **Luigia Petre**

PRECISERING AV SYSTEM

(REFINING SYSTEMS)

When specifying systems, we typically start with a simple model and then we add features onto it until an acceptable level of detail is reached. In order to advance from a simpler model to a more complex one, we need to ensure some properties, for instance, that all the properties of the simpler system are still respected by the more complex one. This 'correct' development is called refinement and is studied in many languages for precisely specifying software and systems. There are many types of refinement, such as algorithmic, data, trace, superposition.

An example of a BSc/MSc thesis that can be written on this topic is, for instance, investigating the similarities and differences of data refinement in CSP and Action Systems. A thesis involves (studying and) describing the refinement concepts, then motivating the chosen type(s) of refinement to be

investigated in the thesis, then the actual content. Larger refinement case studies are also suited well within this topic.

Handledare: **Luigia Petre**

SATISFIABILITY MODULO THEORIES

Satisfiability Modulo Theories (SMT) är problemet att avgöra om en formel i första ordningens logik är satisfierbar under specifika bakgrundsteorier. Typiska bakgrundsteorier är reella talens aritmetik, heltalsaritmetik, och räckmanipulationer. Verktyg för lösning av SMT-problem (SMT solvers) har tillämpningar inom många områden, exempelvis programverifiering, skedulering, optimering.

Handledare: **Johannes Eriksson**

MATEMATIK OCH LOGIK INOM IT

ATT PRESENTERA MATEMATISK TEXT ELEKTRONISKT

Att presentera matematisk text i elektroniskt format har historiskt varit problematiskt. Dessutom kan inmatningen ses som en utmaning, iom att ett vanligt tangentbord endast ger direkt tillgång till ett begränsat antal matematiska symboler. Den studerande skall ge en historisk överblick över de olika system som använts – och som används i dag – både för att presentera och för att mata in matematisk text elektroniskt.

Handledare: **Linda Mannila**

SAMBAND MELLAN KONNEKTIV I SATSLOGIKEN

Man undersöker olika kända och mindre kända konnektiv av olika ställighet och visar på hur de kan uttryckas med varandra.

Handledare: **Patrick Sibelius**

EN KORT HISTORIK ÖVER PREDIKATLOGIKEN

(A SHORT HISTORY OF PREDICATE LOGIC)*(Även möjlig som gradu)*

Man presenterar utvecklingen av predikatlogiken från Aristoteles över Frege och Russell till våra dagar, mera eller mindre detaljerat och djupt beroende på nivån.

Handledare: **Patrick Sibelius**

MOBILA SYSTEM

ANDROID OCH IOS UR ETT UTVECKLARPERSPEKTIV

Den studerande skall ge en introduktion till mobilutveckling samt därefter utvärdera två av plattformarna, Googles Android och Apples iOS, ur ett utvecklarperspektiv. Faktorer som kan tas i beaktande är bl.a. utvecklingsmiljö- och verktyg, applikationsstruktur, distributionskanaler och marknader.

Handledare: **Linda Mannila**

BERÄKNINGAR BEROENDE AV RUM

(SPACE-DEPENDENT COMPUTING)

For several years already, almost everyone has used a GPS to find a new location. This is one of the most visible applications of location-awareness. There are many frameworks that model and analyze various properties for location aware systems.

A thesis can investigate varied issues regarding location-awareness, such as GPS algorithms, location-tracking devices, or more theoretical comparisons of frameworks and systems.

Handledare: **Luigia Petre**

PARALLELLPROGRAMMERING

SCIENTIFIC WORKFLOWS

Handledare: **Mats Aspnäs**

PARALLELL INPUT/OUTPUT

Handledare: **Mats Aspnäs**

PROGRAMMERINGSSPRÅK OCH PROGRAMMERINGSPARADIGMER

SCALA SOM NYBÖRJARSPRÅK I UNDERVISNINGEN

Den studerande skall sätta sig in i litteratur kring de krav som ställs på programmeringsspråk som skall användas i nybörjarundervisningen. Vidare skall den studerande kort presentera programmeringsspråket Scala och analysera dess lämplighet som nybörjarspråk baserat på de krav som funnits i litteraturgenomgången.

Handledare: **Linda Mannila**

PROGRAMMERINGSSPRÅKET <X> OCH ARTIFICIELL INTELLIGENS

(X väljs av skribenten)

Artificiell intelligens är ett fascinerande forskningsområde där man eftersträvar program som klarar av olika sorters uppgifter som allmänt anses kräva intelligens. Ofta är det frågan om att hitta en eller flera lösningar på ett givet problem. För att programmet skulle bete sig intelligent krävs det av programmeraren en speciell satsning på kunskapsrepresentation och olika sökstrategier. Olika språk (ex. Lisp och Prolog) har utvecklats enkom för dessa ändamål, men det är också möjligt att använda ett generellt språk (ex. Java) för att utveckla AI-system. Utgångspunkten för uppsatsen skulle vara att presentera ett av dessa språk och hurdana intelligenta system det kan användas för. Hur stöder just detta språk de strukturer och metoder som artificiell intelligens kräver?

Material (beror på det valda språket, men här finns ett par utgångspunkter):

- Luger, G. & Stubblefield, W.: *AI Algorithms, Data Structures, and Idioms in Prolog, Lisp, and Java*
- Luger, G.: *Artificial Intelligence. Structures and Strategies for Complex Problem Solving*
- Sebesta, R.: *Concepts of Programming Languages*

Handledare: **Annamari Soini**

PROGRAMMERINGSSPRÅKET <X> OCH WEBBPROGRAMMERING

(X väljs av skribenten)

HTML-dokument är i sig helt statiska, men för interaktiva webbsidor krävs en hel del processering. På serversidan möjliggörs detta av CGI (Common Gateway Interface), som tillåter HTML-dokument att begära exekvering av program som finns på servers. Resultaten av dessa beräkningar skickas till webbläsaren i form av HTML-dokument. Processering på klientändan blev möjlig i och med Java applets. Båda dessa tillvägagångssätt håller småningom på att ersättas av nyare teknologier, ofta mha skriptspråk (ex. JavaScript, PHP). Utgångspunkten för uppsatsen skulle vara att presentera ett dylikt språk och hur det används inom webbprogrammering.

Material (beror på det valda språket, men här finns en utgångspunkt):

- Sebesta, R.: *Concepts of Programming Languages*

Handledare: **Annamari Soini**

TYPSYSTEMET I SCALA: TYPÄKERHET VIA TYPINFERENS

Scala är ett modernt multiparadigm programmeringsspråk som förenar objekt-orienterad och funktionell programmering. Scala har statisk typning så som t.ex. Java men för att förenkla koden och göra den mera kompakt används typinferens för att beräkna typer istället för att alltid ge dem explicit. I den här uppsatsen skulle en beskrivning av dels typsystemet (som innehåller flera intressanta koncept) samt metoderna som används för typinferens studeras.

Handledare: **Pontus Boström**

REALTIDSSYSTEM

TIDSBEROENDE BERÄKNINGAR

(TIME-DEPENDENT COMPUTING)

Real-time systems are prevalent in our society. There are numerous time-aware frameworks, either for modelling and analysing systems or for implementing them. Deadlines, deadlocks, and synchronization are only a few of the issues being studied.

A thesis can explore one or several such frameworks for their advantages and disadvantages, can propose solutions, or describe case studies and/or implementations. Comparisons between frameworks are also a good choice for a thesis.

Handledare: **Luigia Petre**

SOCIALA ASPEKTER INOM IT

TEKNISKA MÖJLIGHETER ATT FÖRHINDRA E-MOBBNING

Mobbning som sker med hjälp av IT-teknologi är ett växande problem, speciellt bland barn och ungdomar. Det som skiljer e-mobbning från "vanlig" mobbning t. ex. i skolan är att den mobbade är utsatt praktiskt taget hela tiden; hemmet utgör inte en skyddad sfär när det gäller illasinnade textmeddelanden eller skrivelser i IRC-galleriet och Facebook. Dessutom kan mobbarna vara anonyma så att offret inte längre har någon aning vilka hans eller hennes ovänner är. Det finns ett flertal projekt och initiativ för att förebygga e-mobbning, men de flesta av dessa utgår från attitydfostran. Idén med denna uppsats skulle vara att beskriva e-mobbning som fenomen och undersöka olika tekniska möjligheter för att hindra det eller avslöja personerna bakom det.

Material:

Det finns massor av material om e-mobbning (se e-mobbning som sökord). För tekniska medel mot detta, sök med filter, anonymitet, censur etc. Det tycks inte finnas mycket skrivet om just tekniska medel för att motverka fenomenet, så du får vara kreativ här.

Handledare: **Annamari Soini**

VERKTYG OCH RAMVERK FÖR SYSTEMUTVECKLING

ANIMERINGSVERKTYG SOM HJÄLPMEDEL FÖR ATT SKAPA PÅLITLIGA PROGRAM

Det är viktigt att programvara som styr säkerhetskritiska system fungerar som den är avsedd att fungera. För komplexa system är det dock inte en trivial uppgift. Uppsatsen skulle beskriva hur animeringsverktyg kan användas i systemutvecklingen för att åstadkomma detta. En jämförelse av olika animeringsverktyg för ändamålet kunde göras. Möjliga verktyg är t.ex. ProB Animator eller Brama.

Handledare: **Marina Waldén**

JAVA MODELLING LANGUAGE (JML) OCH DESS STÖDVERKTYG

Java Modelling Language (JML) är ett specifikationspråk för Java som är baserat på programannotationer i form av för- och eftervillkor (kontrakt) och invarianter. JML stöds av ett stort antal verktyg för statisk analys, programbevisning, testgenerering, och debuggning.

Handledare: **Johannes Eriksson**

OSGI-RAMVERKET FÖR JAVA

OSGi är en standard för att dynamiskt installera (även över nätverk), exekvera, och avinstallera Java-applikationer inom en och samma JVM (Java Virtual Machine). Utvecklingsomgivningen Eclipse är ett exempel på en applikation som använder ett OSGi ramverk för dynamisk laddning av insticksprogram (plug-ins).

Handledare: **Johannes Eriksson**

DESIGNMÖNSTRET MODEL-VIEW-CONTROLLER I WEBBAPPLIKATIONER

Model-View-Controller (MVC) är ett grundläggande designmönster inom programvaruutveckling och bygger på separation av en applikation i data (Model), presentation (view), och interaktion (controller). Välj ett specifikt ramverk för webbapplikationsutveckling (t.ex. Struts, Ruby on Rails, Django, Zope) och beskriv hur det implementerar MVC.

Handledare: **Johannes Eriksson**