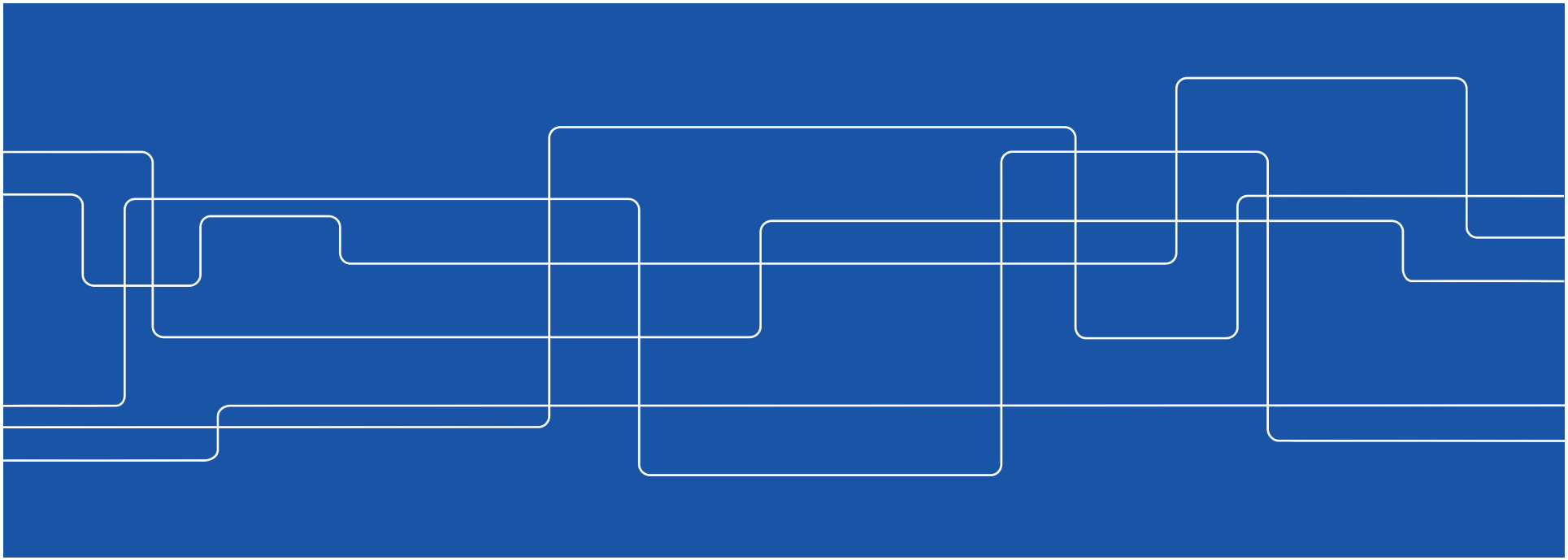
Bättre språk i exjobben –
en väg till ökad synlighet för KTH?

Viggo Kann, Per-Anders Östling,
Greta Quesada Richardson, Gerald Q. Maguire Jr.



KTH:s språkkommitté

- ansvarar för implementeringen av KTH:s språkpolicy i verksamheten
- ger råd i språkliga frågor av generell art

www.kth.se/social/group/sprakkommitten

Språkkommitténs planer 2016

- Informera om språkpolicyn
- Svensk och engelsk terminologi
  KTH:s svensk-engelska ordbok
- Förenkla översättning av texter
- Arbeta för fler språkpersonalkurser
- Lunchseminarium och workshop om språk i exjobben

KTH:s språkpolicy, ur handlingsplanen

4.2 Hög språklig kvalitet

KTH:s främsta språkliga mål, hög språklig kvalitet, ska uppnås som en integrerad del i all utbildning och beaktas fortlöpande i allt arbete.

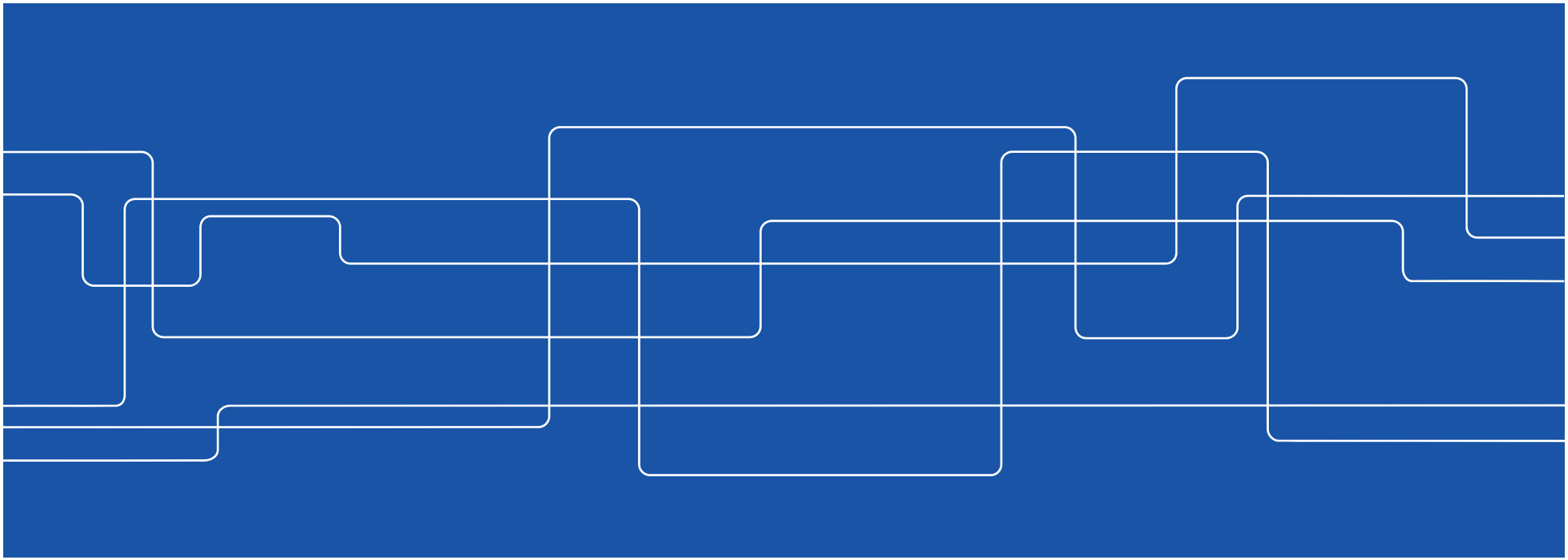
4.10 Sammanfattningar av avhandlingar

Varje examensarbete på grundnivå och avancerad nivå, licentiatavhandling och doktorsavhandling som skrivs på engelska ska ha en svensk sammanfattning.

Synlighet för KTH

-varför är det viktigt?

Per-Anders Östling, planerings- och utredningsavd



Ranking har bl.a. stor inverkan på:

- Politiska beslut och nationella policys
- Rekrytering av internationella studenter och forskare
- Alumners anställningsbarhet
- Konkurrens om externa forskningsmedel
- Samarbete med multinationella företag
- Samarbete med världsledande universitet
- Synlighet och prestige

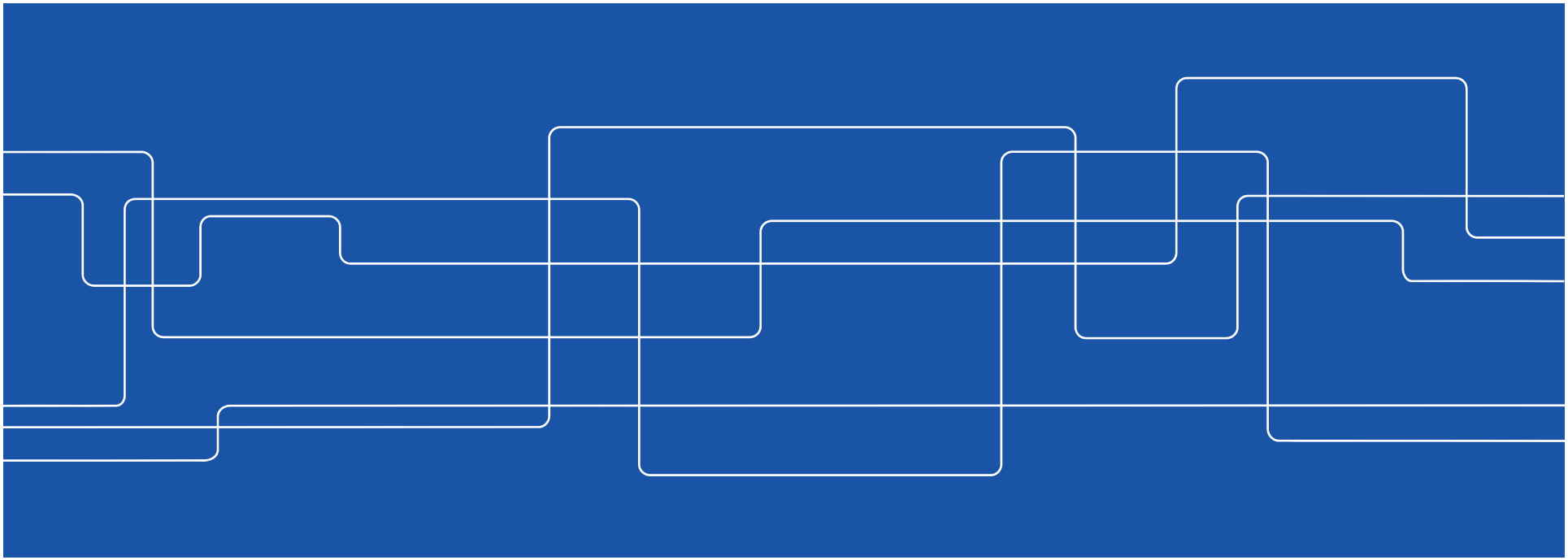
Ökad synlighet och tillgänglighet

- Stärka KTH:s rykte i omvärlden både nationellt och internationellt
- Publicera i synliga och ledande kanaler
- Rätt affiliering
- Sampublicering
- Ökad tillgänglighet av publikationer m.m.
- Publicering i Open Access
- Forskarutbyte, delta i konferenser
- Synlighet på webben
- Stärka varumärket
- Utökad samverkan med näringsliv och omgivande samhälle

Examensarbeten i DiVA

-varför är det viktigt?

Greta Quesada Richardson, publiceringens infrastruktur



KTH:s publikationsdatabas DiVA

Studentuppsatser i DiVA

- Förvaltningsbeslut https://www.kth.se/polopoly_fs/1.589177!/UF20100650.pdf omfattar alla studentuppsatser från 2011 och framåt
- Alla godkända studentuppsatser ska registreras i DiVA
- En PDF med uppsatsen i fulltext ska publiceras i DiVA, om inga upphovsrätt sätter hinder – studenten godkänner detta

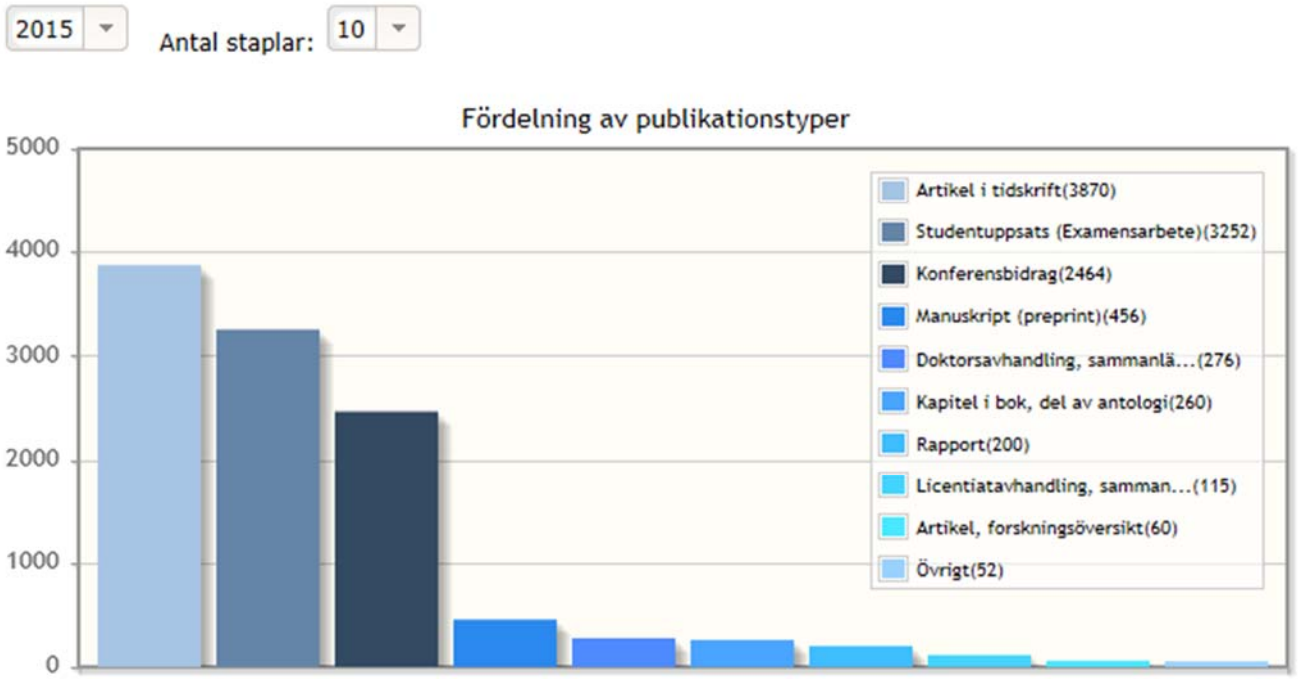
Studentuppsatser finns bara i DiVA

- **Tillgänglighet**. Biblioteket tar inte emot tryckta examensarbeten, finns bara i DiVA.

- **Synliggöra**. Examensarbeten i DiVA blir sökbara även i andra databaser exempelvis i Uppsök, Google och Google Scholar.

- **Spridning och citering.** Publikationer på nätet citeras mer än det som bara finns i tryckt form.

Fördelning mellan examensarbeten och forskningspublikationer



Examensarbeten får stor spridning, år 2015

r-Process Simulation and Heavy-Element Nucleosynthesis.

Alvelid, Jonatan

KTH, Skolan för teknikvetenskap (SCI), Fysik.

2014 (Engelska)

Självständigt arbete på grundnivå (kandidatexamen), 10 poäng / 15 hp

Studentuppsats (Examensarbete)

Abstract [en]

r-process, short for rapid neutron capture process, is a nucleosynthesis process taking place on short time scales. Rapid neutron captures produce less and less stable neutron-rich nuclei which in turn beta minus decays when the probability for beta decay is higher than the probability for neutron captures, upon which more neutrons are captured and the process repeats itself, creating r-process paths. Very neutron-rich heavy elements are the product of this process taking place at explosive astrophysical sites with high neutron flux. Simulations of r-processes are important for finding out the exact sites, something that is yet not known. To get more accurate simulation results leading to a better understanding of r-processes, the initial parameter dependence of the simulations is important to understand. This report discusses the dependence on three important initial parameters; temperature, density and electron fraction. Furthermore, the dependence on nuclear masses is covered, which is important since no exact model for nuclear masses exists for the neutron-rich nuclei involved. Finally, different stopping criteria are simulated, representing different physical environments in which r-processes may occur. Results from the simulations, carried out using r-Java 2.0, show that r-process simulations are sensitive to all parameters discussed; further research can tell to which extent. A better understanding of the dependence on the parameters will hopefully extend our knowledge of r-processes and where in the universe they occur.

Abstract [sv]

r-process, rapid neutron capture process, är en snabb nukleosyntesprocess. Snabba neutroninfångningar producerar allt mer instabila neutronrika atomkärnor som slutligen betaminussönderfaller när sannolikheten för betasönderfall blir högre än sannolikheten för en ny neutroninfångning. Därefeter fångas fler neutroner och processen upprepar sig själv i r-processkedjor. Väldigt neutronrika tunga ämnen bildas under denna process som kräver explosiva astrofysikaliska platser med höga neutronflux. Då det ännu är okänt exakt var dessa platser är så hjälper r-processsimulationer att förstå detta. För att förbättra simuleringsresultaten och därmed förståelsen av r-processer så är det viktigt att förstå hur initiala parametrar påverkar simuleringarna. Temperatur, densitet och förhållandet mellan fria elektroner och nukleoner är tre parametrar som denna rapport behandlar. Påverkan av kärrnmassor diskuteras också, vilket är viktigt då ingen exakt modell för kärnmassor existerar. Slutligen behandlas även olika stoppkriterium vilket representerar olika fysikaliska miljöer där r-processer eventuellt förekommer. Resultat från simuleringar, gjorda i r-Java 2.0, visar på att r-processimuleringar är känsliga för alla parametrar som har behandlats men där vidare forskning får visa till vilken grad. En bättre förståelse för hur simuleringarna påverkas av parametrar kommer förhoppningsvis öka förståelsen för r-processer och var i universum de förekommer.

Ort, förlag, år, upplaga, sidor

2014. , 25 s.

Mest nedladdade avhandlingen år 2015

On design and analysis of synchronous permanent magnet machines for field-weakening operation in hybrid electric vehicles

Magnussen, Freddy

KTH, Tidigare Institutioner, Elektrotekniska system.

2004 (Engelska)

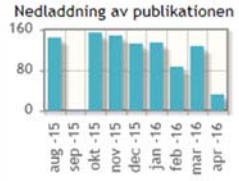
Doktorsavhandling, sammanläggning (Övrigt vetenskapligt)

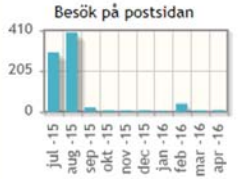
Abstract [en]

A regular vehicle of today is equipped with an internal combustion engine that runs on either gasoline or diesel, which are fossil fuels from oil reserves that are millions of years old. In all types of combustion processes carbon dioxide and several other emissions are produced. There are none known technologies of today that can reduce the emissions of carbon dioxide from combustion, but the amount that is produced is mainly dependent on the fuel that is used. Combustion of fossil fuels increases the contamination of carbon dioxide in the atmosphere and diminishes the oil resources. The results are global warming and empty oil reserves within a few decades with the current production tempo, in addition to many other pollution effects that are harmful to the environment. A transition towards a society based on sustainable transportation is therefore urgent. The hydrogen fuel cell powered car with an electric propulsion system has the potential to be the car of the future that possesses the required characteristics of no harmful tailpipe emissions. There are some obstacles in the way for an early commercialisation, including the expensive catalysts used today and the lack of an infrastructure based on hydrogen, though. The hybrid electric vehicle, with both a conventional as well as an electric drivetrain, is a natural candidate for making the transition from the conventional car towards the car of the future.

This thesis is focused on the design and analysis of permanent magnet machines for a novel hybrid electric vehicle drive system called the Four Quadrant Transducer. A number of electrical machine aspects are identified, including cores of soft magnetic composites, fractional pitch concentrated windings, core segmentation, novel machine topologies and cost effective production methods. The main objective is to analyse and judge the many unconventional machine aspects of which some may have the potential to improve the performance and reduce the cost of permanent magnet machines. Another objective is to study the effects of the use of fossil fuels and describe them with a new perspective and thereby make one small contribution to the debate about energy issues. Much focus has been spent on the theory of concentrated windings for permanent magnet machines. The potential parasitic effects and methods to improve the torque performance have been described. Other topics that have been given a high priority are material and power loss studies. An important contribution to the understanding of iron losses during field-weakening operation has been presented. A comprehensive use of finite element modeling has been done in the analysis combined with measurements on several laboratory prototypes.

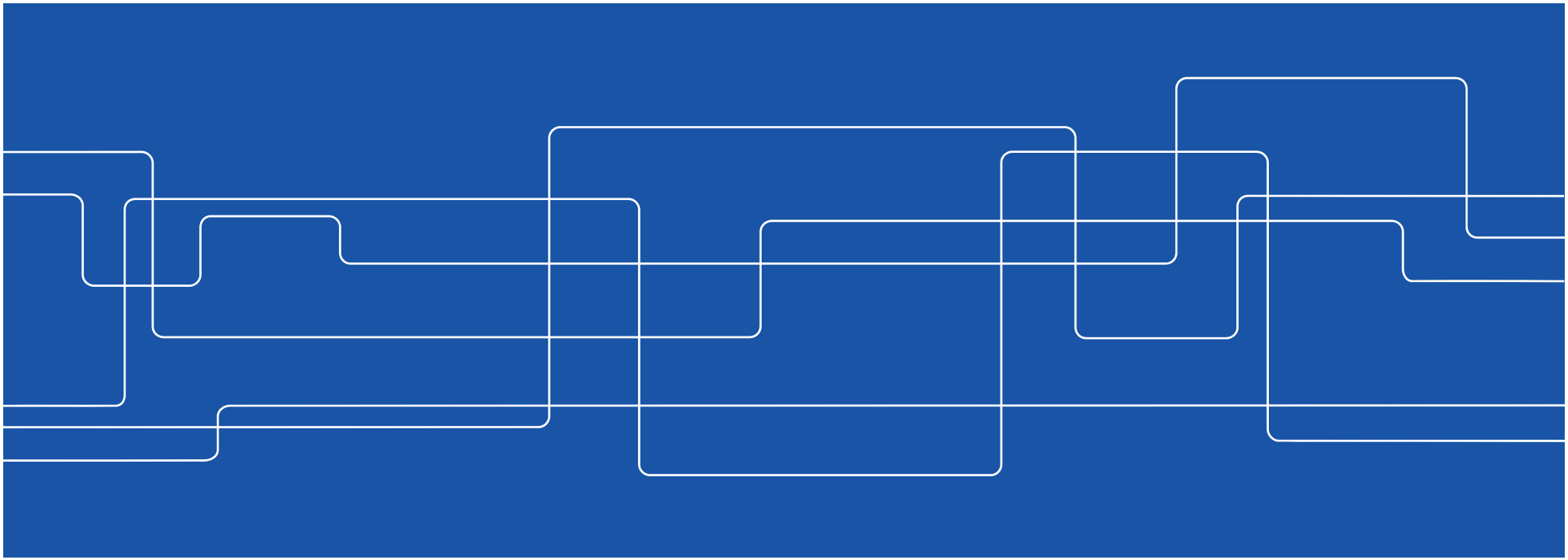
The Four Quadrant Transducer drivetrain and its two electrical machines intended for a midsized passenger car has been studied. The gearbox can be of a simple single stage type, which reduces the mechanical complexity and makes the traction performance of the vehicle smooth, without gear changes and drops in power. Simulations on a complete hybrid system show that fuel savings of more than 40% compared to a conventional vehicle can be achieved at citytraffic driving. The savings are modest at





Undersökning av språk i exjobben

Gerald Q. Maguire Jr., språkkommittén och ICT



Motivation

“Language Policy for KTH 2010”, Dnr V-2009-0365, doss 1, section 4.10 Abstracts of dissertations and theses "Every Master’s dissertation, licenciate thesis and doctoral thesis that is written in English is to include a Swedish Abstract. Similarly, every dissertation or thesis written in Swedish is to have an English Abstract."

http://intra.kth.se/en/regelverk/policyer/sprakpolicy-for-kth-2010-1.60909

All theses are to be reported in KTH’s Digitala Vetenskapliga Arkivet (DiVA) since 1 January 2011 according to “**Elektronisk registrering av examensarbeten i DiVA**” **Beslutsnummer:** UF-2010/0650

⇒ Hence it should be possible to mechanically determine if the language policy has been followed

Goals

- **Making theses more visible:** in 2015: 2256 theses and 4886 research publications at KTH (of which 412 were licentiate and doctoral theses/dissertations)
- Improve abstracts and keywords for each thesis
- Meeting goals of the “Language Policy for KTH 2010”, Swedish law, and addressing the needs of both English and Swedish speaking students in ICT (especially due to the introduction of a 3 year program taught in English)

$\Rightarrow$ Increased student **visibility** and improved thesis **quality**

Timely visibility of theses

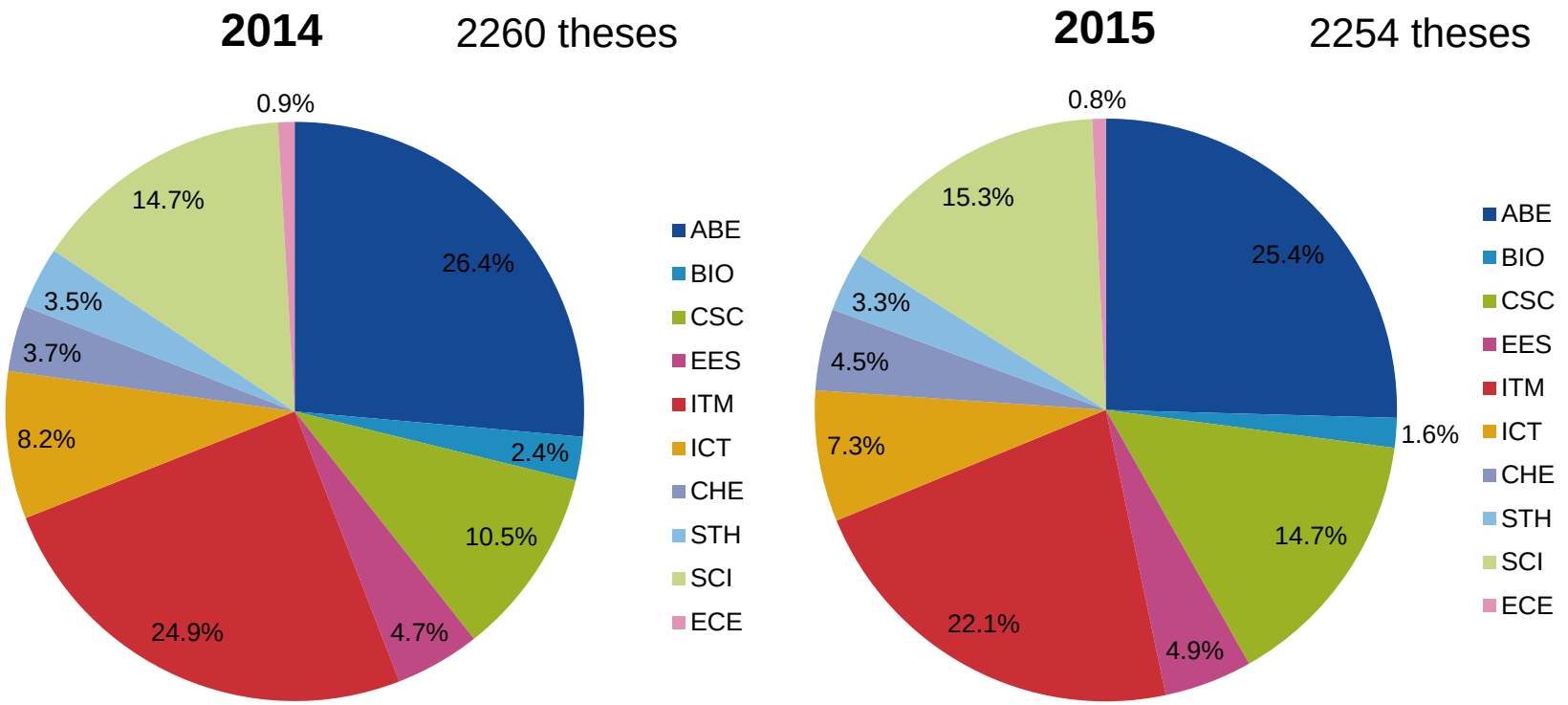
In 2015, I had two instances of theses being sought as prior work with respect to patents:

1. In a US court case
2. By the European Patent Office

A critical issue was: **When** did the document become publically available?

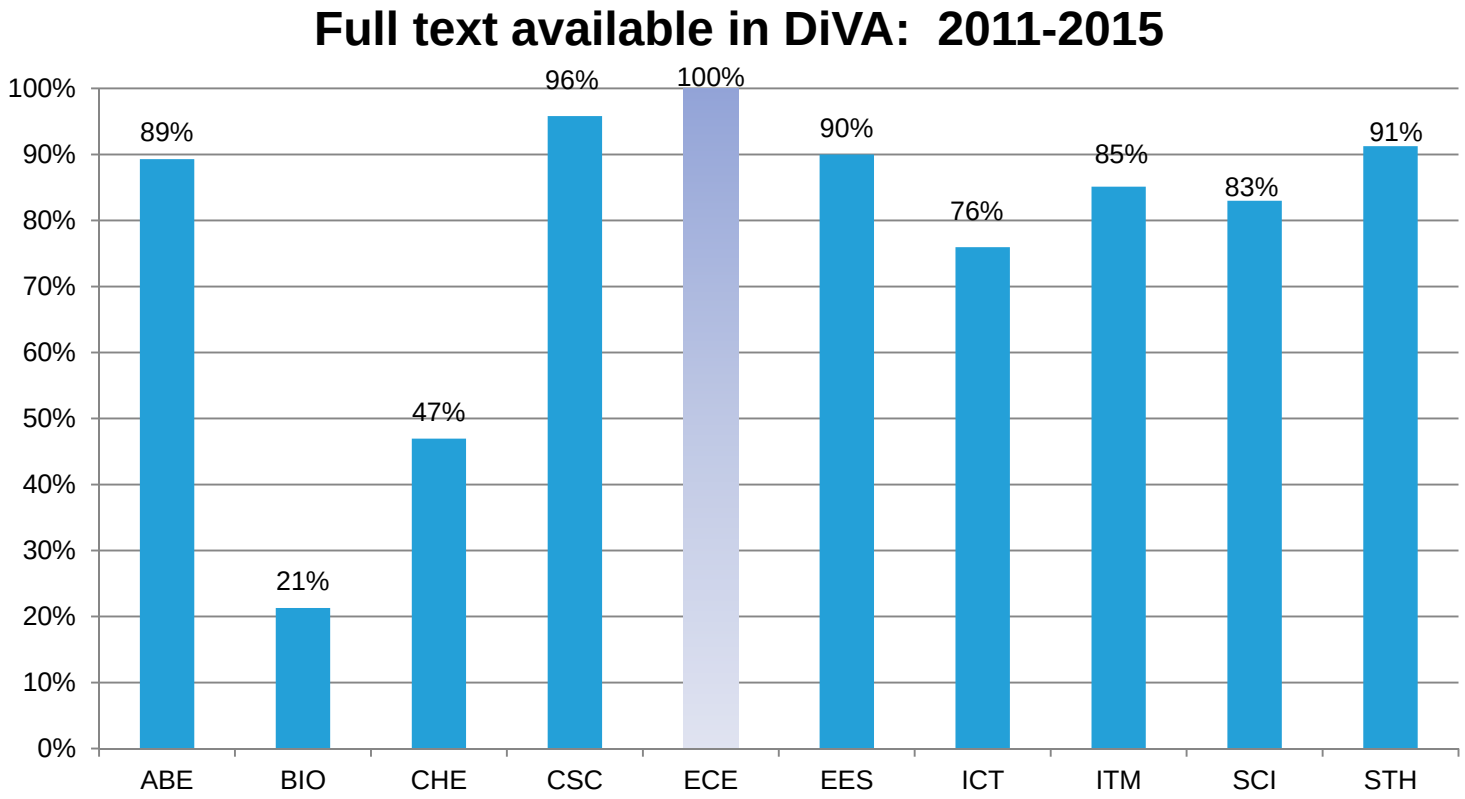
⇒ Faculty and administrators **must** process theses in a timely manner. (Note that formally theses are subject to "Offentlighetsprincipen".)

Distribution of student theses over schools (based on DiVA entries)

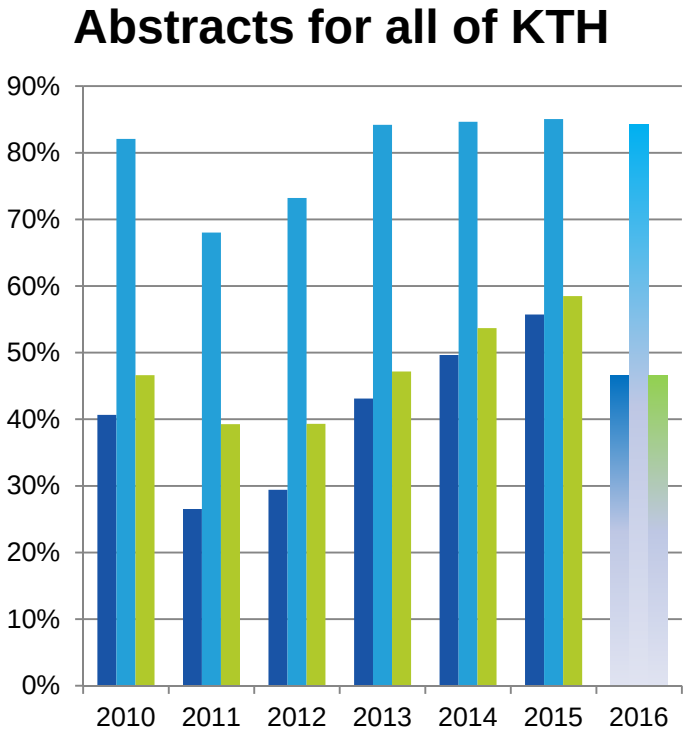


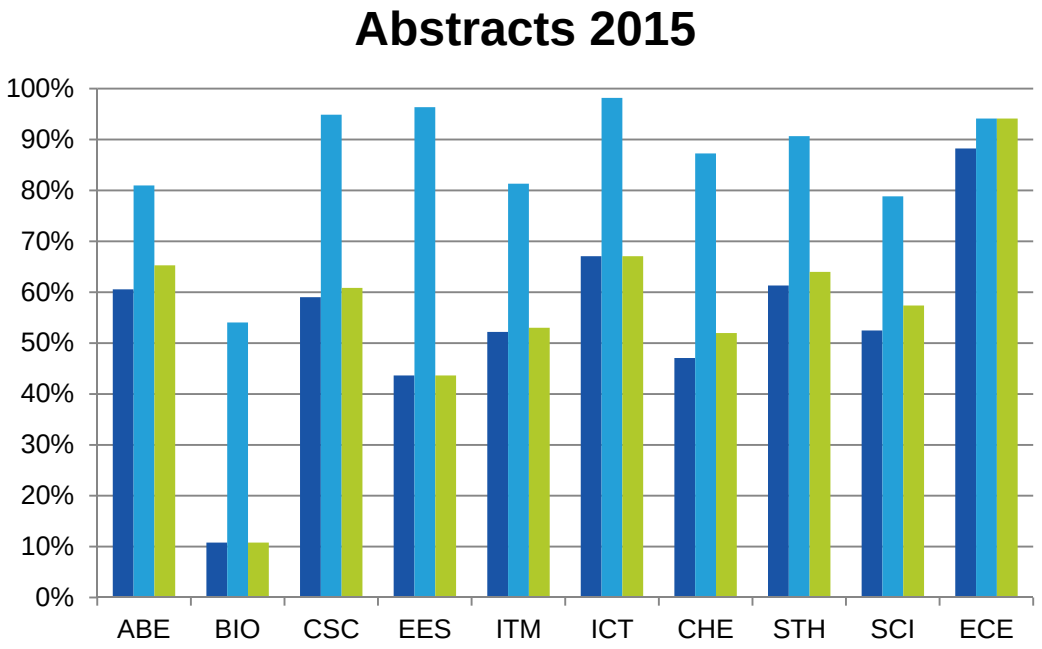
Note that number of theses is **not** the same as the number of students completing theses, as some theses have multiple authors. Year is year of publication (not the year in which the degree projected was completed.

Full text is generally available: (except for BIO and CHE)



Abstracts in English, Swedish, and both





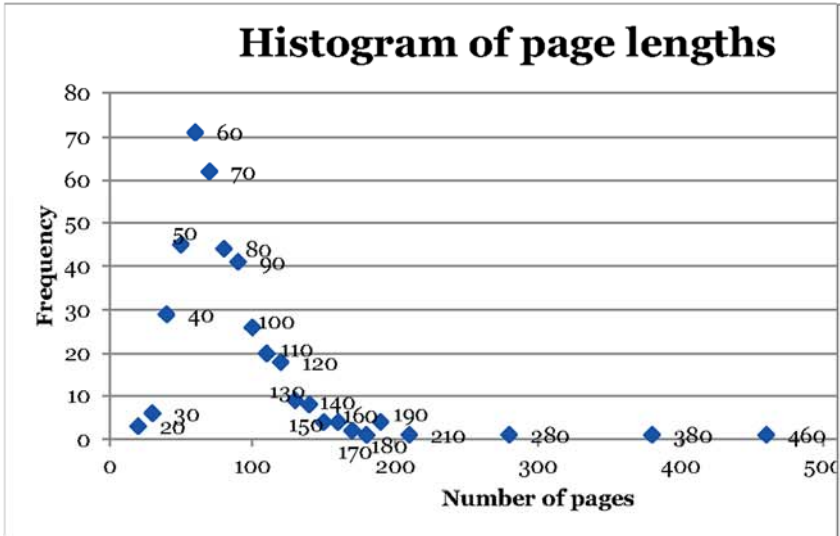
Language of abstracts in DiVA: 2015

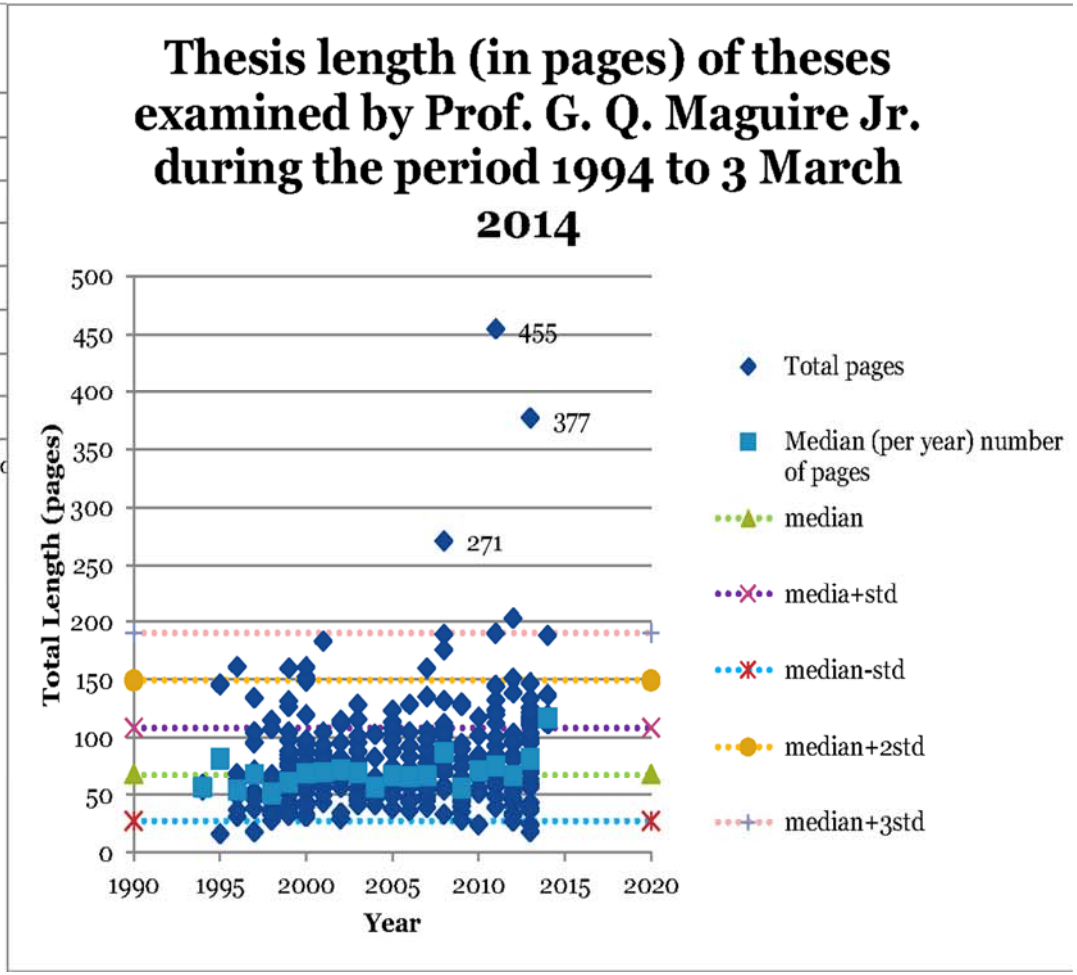
School	count	Fraction of theses	Abstracts_eng_swe		Abstracts_eng		Abstracts_swe		Abstracts_missing		Abstracts_nor	Abstracts_ger
ABE	573	25.4%	347	60.6%	464	81.0%	374	65.3%	81	14.1%	1	
BIO	37	1.6%	4	10.8%	20	54.1%	4	10.8%	17	45.9%		1
CSC	332	14.7%	196	59.0%	315	94.9%	202	60.8%	11	3.3%		
EES	110	4.9%	48	43.6%	106	96.4%	48	43.6%	4	3.6%		
ITM	498	22.1%	260	52.2%	405	81.3%	264	53.0%	89	17.9%		
ICT	164	7.3%	110	67.1%	161	98.2%	110	67.1%	3	1.8%		
CHE	102	4.5%	48	47.1%	89	87.3%	53	52.0%	8	7.8%		
STH	75	3.3%	46	61.3%	68	90.7%	48	64.0%	5	6.7%		
SCI	345	15.3%	181	52.5%	272	78.8%	198	57.4%	56	16.2%		3
ECE	17	0.8%	15	88.2%	16	94.1%	16	94.1%	0	0.0%		
Old inst.	1	0.0%	1	100.0%	1	100.0%	1	100.0%	0	0.0%		
Total	2254	100.0%	1256	55.7%	1917	85.0%	1318	58.5%	274	12.2%		

Other uses of the data from DiVA

- Statistics
  - Length of theses
  - Lengths of abstracts
- Keywords
  - Collected keywords
  - Pairing English and Swedish keywords
- Generating vocabulary lists
- Improved searching
- Information about citations

Length of theses – “How long is a thesis?”





Abstract: structure

- **What** is the topic area? (optional)
- Short **problem statement**
- **Why** was this problem worth an Kand./M.Sc. thesis project? (i.e., why is the problem both significant and of suitable degree of difficulty to be worth a degree project? Why has no one else solved it yet?)
- **How** did you solve the problem? What was your method/insight?
- **Results/Conclusions/Consequences/Impact**: What are your key results/conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?

Abstracts in more detail (2010-June 2014)

		Abstract					English Abstract			Swedish Abstract		
School	#theses since 2010	English	Swedish	English only	Swedish only	Both	Median (words)	Min	Max	Median (words)	Min	Max
ABE	2122	67%	56%	32%	21%	35%	250	5%	1465	203	1	1530
BIO	67	13%	4%	9%	0%	4%	276	112	513	293	187	442
CHE	168	81%	15%	69%	4%	12%	259	64	674	301.5	51	451
CSC	694	97%	89%	9%	1%	89%	173	55	564	155	51	490
ECE	13	92%	77%	15%	0%	77%	288	104	450	225.5	124	367
EES	346	90%	3%	87%	0%	3%	226	45	704	228	89	637
ICT	792	95%	33%	66%	5%	29%	241	62	644	207	61	572
ITM	1886	78%	40%	41%	3%	37%	286	1	1465	271	1	1530
SCI	1011	23%	7%	12%	2%	3%	204	5	1111	209	41	778
STH	135	51%	50%	15%	14%	36%	258	42	552	229	24	465
(median)	520	80%	36%	24%	2%	32%	254	50	659	227	51	531

Length of abstracts

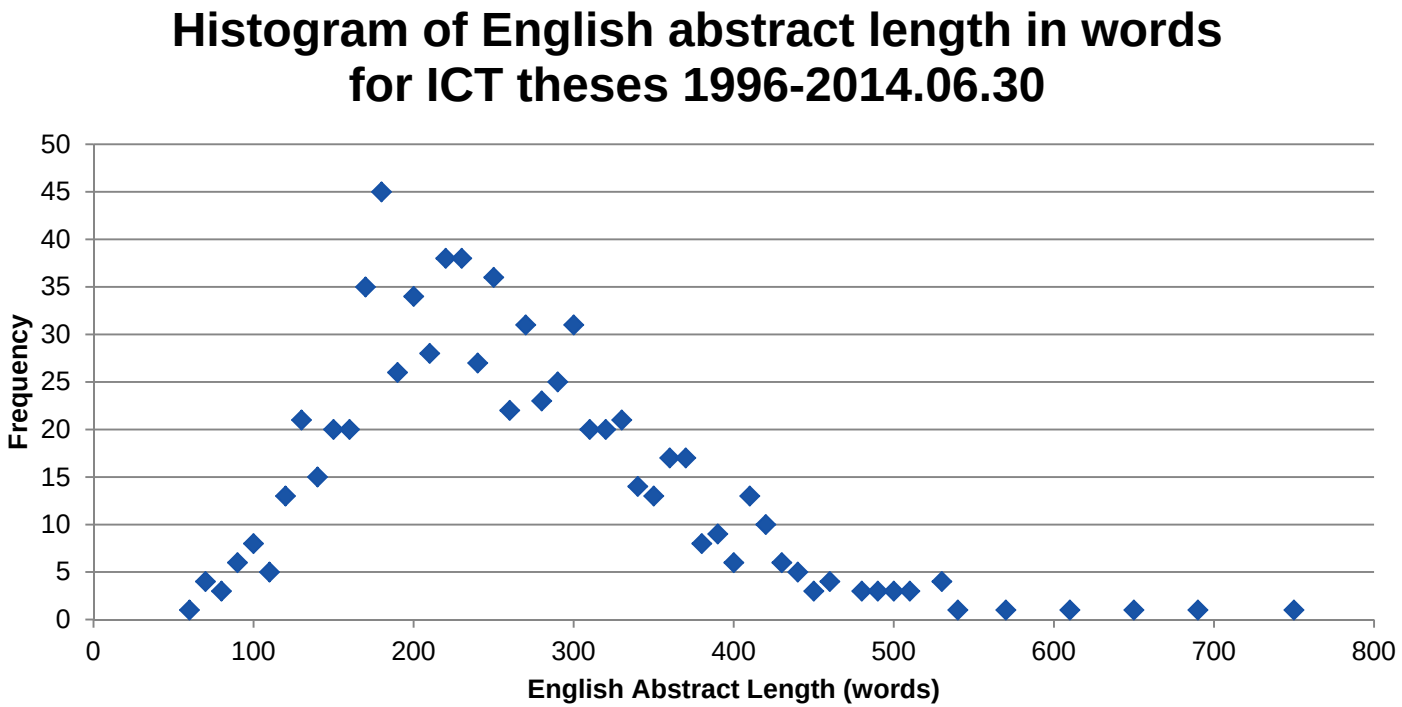
All examiners at ICT		**Maguire**		
in English	in Swedish	in English	in Swedish	description
241	**207**	**289**	**207**	median
62	61	100	98	minimum (non-zero)
644	572	606	509	maximum

At ICT, the median length is 3 paragraphs with a standard deviation of 1.9 paragraphs.

~90% of all English language abstracts for theses examined by prof. Maguire are between 200 and 450 words.
Chalmers University of Technology suggests a Master's thesis abstract should be between 250 and 350 words.
Databases such as ProQuest Dissertations & Theses (PQDT) limit abstracts to 350 words. While Canada's National Archive limits Masters thesis abstracts to 150 words (and Doctoral dissertation to 350 words).

I have not presented any of the statistics concerning abstracts in other languages, e.g. Spanish, French, Italian, Chinese, …
Many double degree students need to have an abstract in the *language of their home university*.

Histogram of number of words in ICT abstracts



Keywords

Choosing good keywords can help others to locate the thesis and related theses.

Choose the most specific keyword from those used in your domain, see for example:

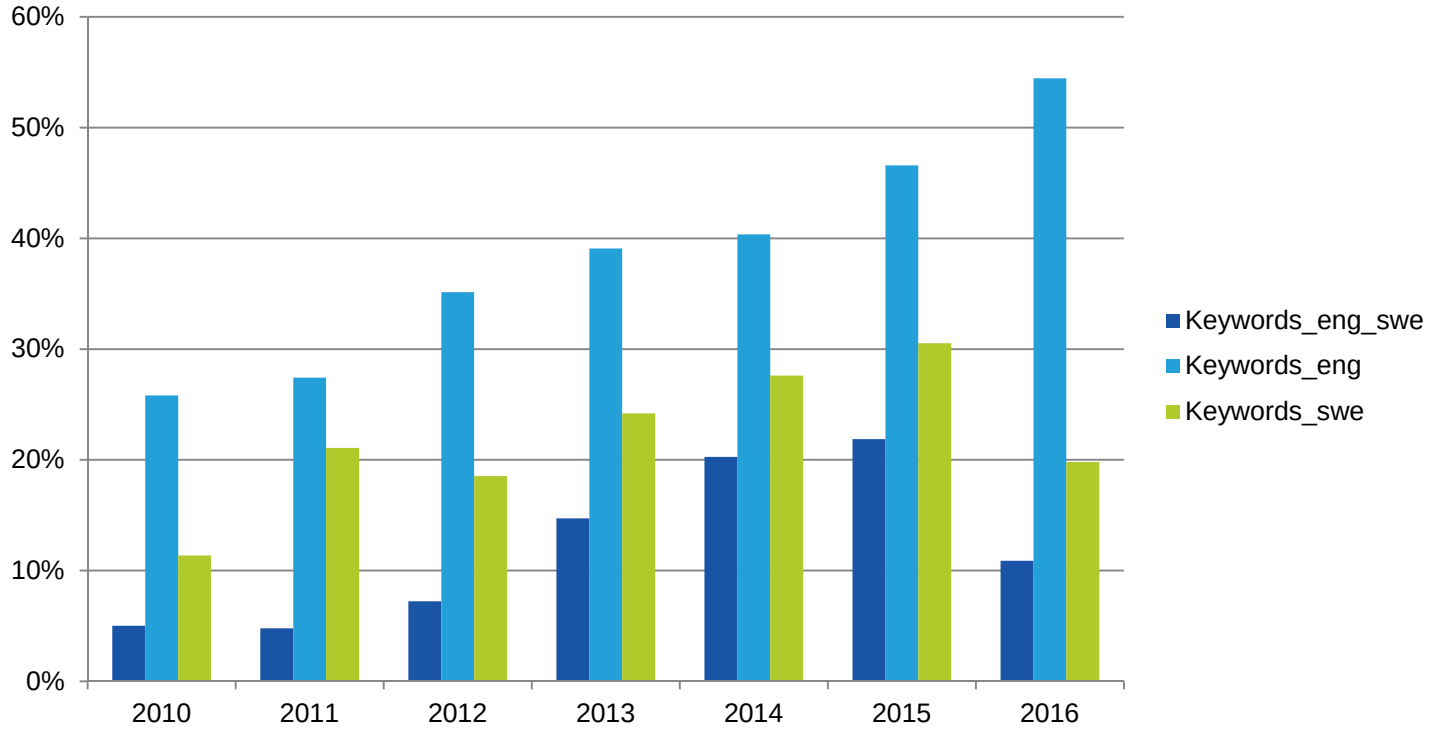
- ACM's Computing Classification System (2012)
- 2013 IEEE Taxonomy Version 1.0

Mechanics:

- The first letter of a keyword should be set with a capital letter and proper names should be capitalized as usual.
- Spell out acronyms and abbreviations.
- Avoid "stop words" - as they generally carry little or no information.
- List your keywords separated by commas (",").

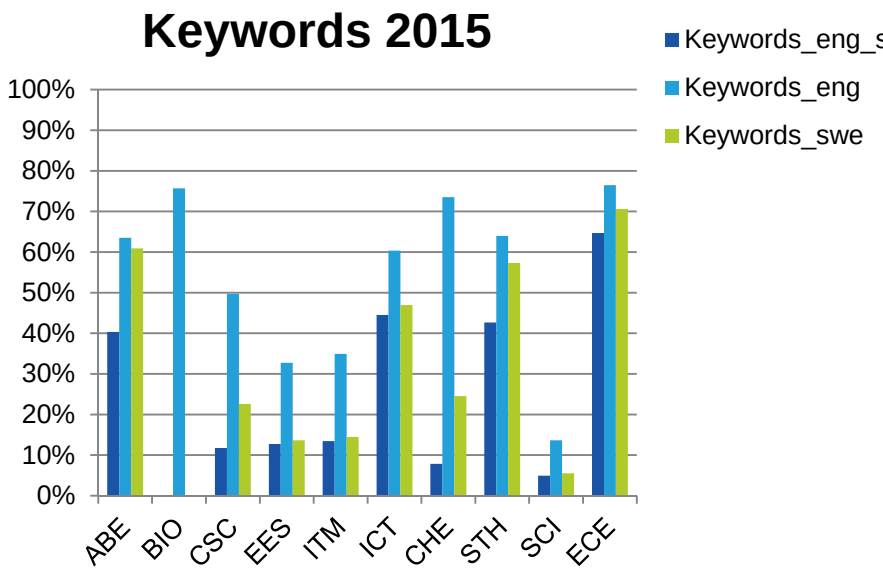
Since you should have both English and Swedish keywords - you might think of ordering them in corresponding order (i.e., so that the n^{th} word in each list correspond) - thus it would be easier to mechanically find matching keywords.

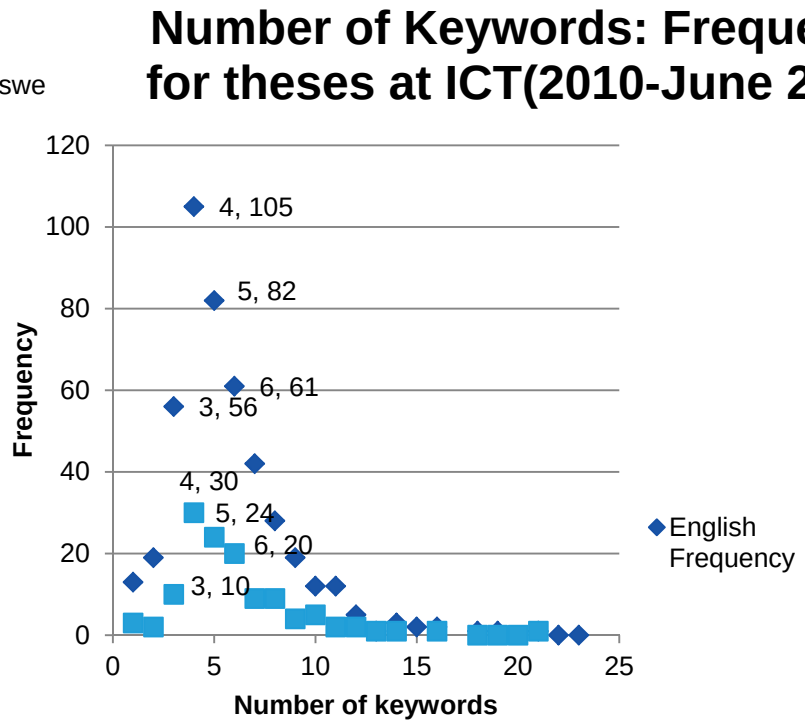
Use of keywords across KTH is increasing



Growth in having English keywords ~4.6% per year, while having Swedish **and** English keywords is growing about ~2.3% per year.

Keywords





3 to 6 keywords are common

Collected keywords: most common

ICT	EES	CSC	ITM	ABE	SCI	CHE	STH	ECE
system|s	system|s	system|s	system|s	building|s	model|s	material|s	work|s	school|s
network|s	model|s	method|s	energy	energy	method|s	water|s	system|s	student|s
service|s	power|s	user|s	work|s	system|s	system|s	sample|s	service|s	teachers
user|s	control|s	data	model|s	area|s	data	process	method|s	management
application|s	network|s	algorithm|s	development|s	method|s	simulation|s	method|s	data	ICT
data	method|s	model|s	product|s	environmental	order|s	properties	level|s	program|s
mobile|s	data	work|s	method|s	construction|s	time|s	showed	patient|s	waste
solution|s	simulation|s	time|s	process|s	work|s	work|s	temperature|s	environment|s	work
device|s	time|s	application|s	companies	model|s	part|s	system|s	device|s	tool|s
design|s	performance|s	information	company|s	development|s	design|s	effect|s	risk|s	development
time|s	algorithm|s	report|s	production|s	process|s	test|s	amount|s	factor|s	engineer|s
performance|s	order|s	game|s	order|s	time|s	material|s	time|s	time|s	competence
information	work|s	order|s	management|s	water|s	risk|s	production	application|s	education
model|s	application|s	solution|s	time|s	design|s	structure|s	reaction|s	model|s	interviews
work|s	sensor|s	test|s	market|s	Stockholm|'s	parameter|s	work|s	part|s	lean
communication|s	information	design|s	cost|s	cost|s	vehicle|s	c	ultrasound	Skanska
method|s	communication|s	development|s	environmental	part|s	case|s	concentration|s	order	learning
order|s	grid|s	media|s	knowledge	material|s	flow|s	oil|s	image|s	meeting|s
test|s	part|s	social	part|s	case|s	algorithm|s	pulp|s	information	part|s

…

x|s indicates the strings x and xs

Generating vocabulary lists

English - Word or Words	Swedish
1 pulse per second	1 puls per sekund
2D method	2D-metod
2G (Third Generation)	2G (andra generationen)
3D (three dimensional)	tre dimensionell
…	
3D-printing	3D-skrivare

...

interest interests interesting interestingly	intressanta intressen intressant intressant
interface interfaces interfaced interfacing	gränssnitt gränssnitten gränssnitts gränssnitt
interface to the router system	gränssnitt till routern systemet
interfere interference	interferera störningar
intergalactic	intergalaktiska
inter-ISP traffic	inter-ISP trafik
interleaving	interfoliering
intermediate intermediary	mellanliggande mellanhand
internal	inre/intern
international	internationell
internet internetworked internetworking	net "av nätverk" internet
internet connectivity	internet-anslutning
internet of things (IoT)	Sakernas Internet
Internet Protocol (IP)	
internet service provider (ISP)	Internetleverantör
internet traffic	internettrafik
internet-draft internet-drafts	Internet-utkast
internetwork internetworking internetworked	internet Internetteknologi/samverkan "sammankopplade nätverk"

...

	zero	noll
	zero knowledge	noll kunskap
3854	zone|s	zon zoner

Desired outcomes

- On-line module for how to write abstracts and selecting keywords
- Paired vocabulary (⇒ policy for ordering of keywords)
- Improved set of stop words – to produce more relevant and useful lists of keywords
- Clustering/linking between theses – to improve searching for related work
  - Clicking on a keyword in a DiVA page for a thesis ⇒ a list of theses that all use this keyword.
  - List of keywords and frequencies could enhance knowing what terms one should use when searching. For example, “synchronization” appears 79 times in abstracts for ICT and is a keyword for 7 theses, 11 times for EES, …
- More tools to help users
  - Use and find existing information
  - To *improve* the quality of the meta data that is kept – for example, mechanical checks for missing spaces, wrong separators, etc.
  - To mechanically produce additional meta data – for example the plots shown in the earlier slides, statistics about thesis abstracts & keywords, …
  - Tools to minimize human effort while maximizing the correct entry of data

Open Questions

- Should each thesis have a DOI, rather than simply a TRITA number?
- Can DiVA's export function for theses be improved to create a more complete reference?
  ⇒ Increasing them being cited
- If thesis reference lists contain DOIs, ISBNs, URLS, can we mechanically create lists of references and lists of citations for each thesis - as is done for example in the ACM Digital Library or WoS?

  If so, what is the format for this data and how can it be added to the DiVA record?

¿Questions?

Abstracts:

- Why are many students and examiners **not** following the “Language Policy for KTH 2010”?
- Why should we encourage students to follow the policy?
- Should we “enforce” the policy?
- What resources are available to help make the process simpler?

Keywords:

- Should we request/require them?
- What resources are available to help make the process simpler?

Code

https://gitr.sys.kth.se/maguire/DiVA-tools

https://gitr.sys.kth.se/maguire/DiVA-tools-new

Are the numbers correct?
Closer look at 2014
⇒ clear need for quality control

In VIS (LADOK) Number of entries for grades		DiVA entries	In LADOK but not in DiVA	in LADOK but missing with 2 authors	Actual missing theses	In DiVA but not in LADOK	Duplicate in DiVA	DiVA thesis is in 2013, but grades is in 2014 or 2015	in DiVA under co-author	Entries in LADOK that do not correspond to a thesis
3	Externa högskolor									
996	Skolan för arkitektur och samhällsbyggnad (ABE)	596								
103	Skolan för bioteknologi (BIO)	55								
495	Skolan för datavetenskap och kommunikation (CSC)	238	64		40 (MSc)+ missing BSc theses	13	6	4	80	128
174	Skolan för elektro- och systemteknik (EES)	107	76			5	2	3		
995	Skolan för industriell teknik och management (ITM)	562	505			65	5	17		
280	Skolan för info.s- och kommunikationsteknik (ICT)	186	70	8	54	18	6	48	34	
132	Skolan för kemivetenskap (CHE)	83								
139	Skolan för teknik och hälsa (STH)	78								
465	Skolan för teknikvetenskap (SCI)	332								
24	Skolan för teknikvet.kommunikation o lärande(ECE)	21	7			4			3	

Workshop om språk i exjobben

26 maj kl 12.30-14.30 i Salongen, KTHB

- Civilingenjörsstudenter gör två exjobb och ska visa kompetens i både svenska och engelska.
  Är det då lämpligt att låta kandidatexjobbet skrivas på svenska och masterexjobbet på engelska?
- Ska den svenska sammanfattningen i exjobben vara en mer populär beskrivning av arbetet än det engelska abstract? Ska exjobbaren erbjudas professionell hjälp med att skriva sammanfattningen?
- Hur kan vi hjälpa exjobbaren att välja bra nyckelord på engelska och svenska?
- Hur kan vi höja (lägsta)nivån på den språkliga kvaliteten i exjobben?

Alla KTH-anställda är välkomna! En lätt lunch serveras.

Anmälan görs på språkkommitténs gruppwebb
www.kth.se/social/group/sprakkommitten