

Språkteknologi

Denna kurs handlar om **naturliga språk** (svenska, engelska, japanska, arabiska...), och ...

... hur vi kan få datorer att utföra **användbara** och **intressanta** uppgifter med naturliga språk.

Språkteknologi

... **användbara** och **intressanta** uppgifter med naturliga språk, till exempel:

- **Verktyg** som underlättar vid skrivning, läsning, informationssökning, sammanfattning, översättning, mm.
- **Gränssnitt** i naturligt språk.

Språkteknologi

- **Fråga 1:** Hur får vi datorer att bearbeta mänskligt språk och utföra intressanta uppgifter?
 - Formell representation av (någon aspekt) av mänskligt språk
 - Algoritmer för att manipulera dessa representationer
- **Fråga 2:** Vad kan dessa formella representationer och algoritmer lära oss om mänskligt språk?

Språkteknologi

- **Tvåvetenskapligt ämne:** datavetenskap, lingvistik, psykologi, matematik, statistik, signalbehandling
- Relaterade termer:
 - NLP (Natural Language Processing)
 - Datalingvistik

Varför är språkteknologi viktigt?

- Enormt mycket kunskap finns tillgängligt som löpande text.
- En stor del av kommunikationen mellan människor går via datorer (chat, mail, Facebook, bloggar, telefonisamtal, ...)
- Naturligt-språk-gränssnitt blir allt vanligare.
- Stort kommersiellt intresse (Google, ...)

The screenshot shows a Google search interface with the query 'språkteknologi kth'. Below the search bar, there are navigation links for 'Webben', 'Kartor', 'Bilder', 'Videor', 'Nyheter', 'Fler', and 'Sökverktyg'. The search results are as follows:

Ungefär 4 730 resultat (0,29 sekunder)

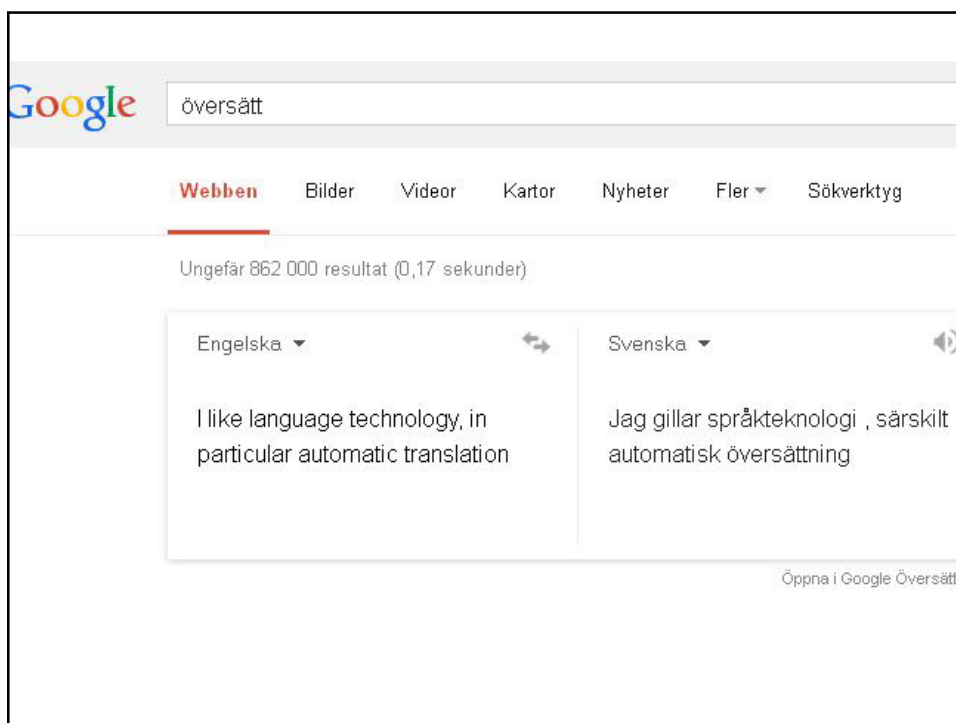
KTH | DH2418 Språkteknologi 6,0 hp
www.kth.se/student/kurser/kurs/DH2418
 Kursen i **språkteknologi** behandlar olika metoder för analys, generering och filtrering av mänskligt språk med fokus på text. Regelbaserade och statistiska ...

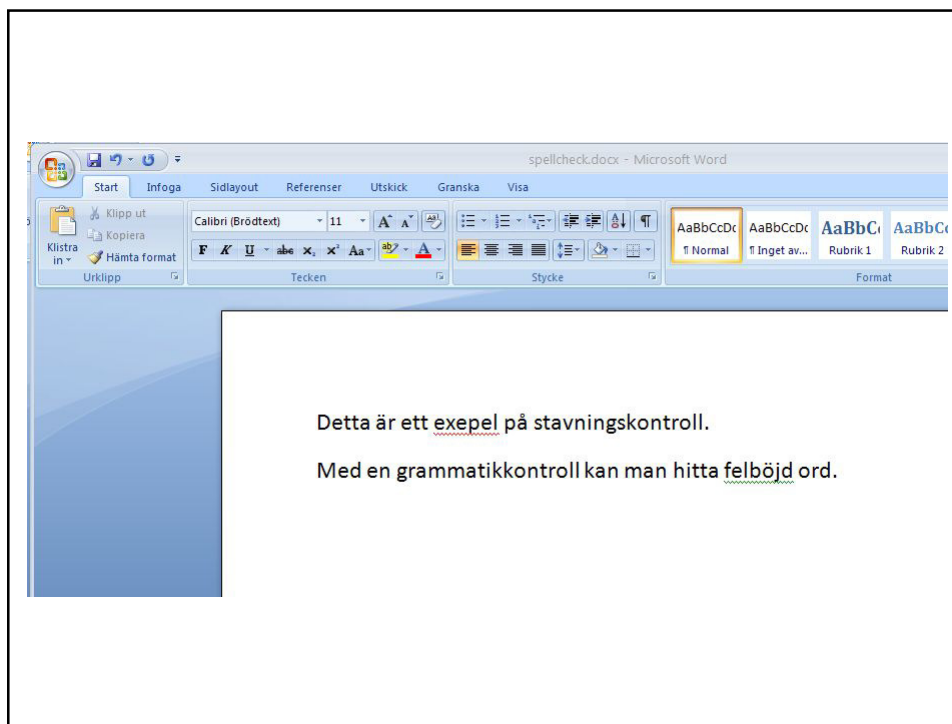
KTH CSC sprakt13
www.csc.kth.se/CSC/Kurser/DD2418
Språkteknologi, sprakt13. Kursomgången är avslutad. Aktuell kursomgång beskrivs på kurswebben. Om du ändå vill titta på den gamla kurssidan för sprakt13 så ...
 Du har besökt den här sidan många gånger. Sidan besöktes senast: 2014-10-20

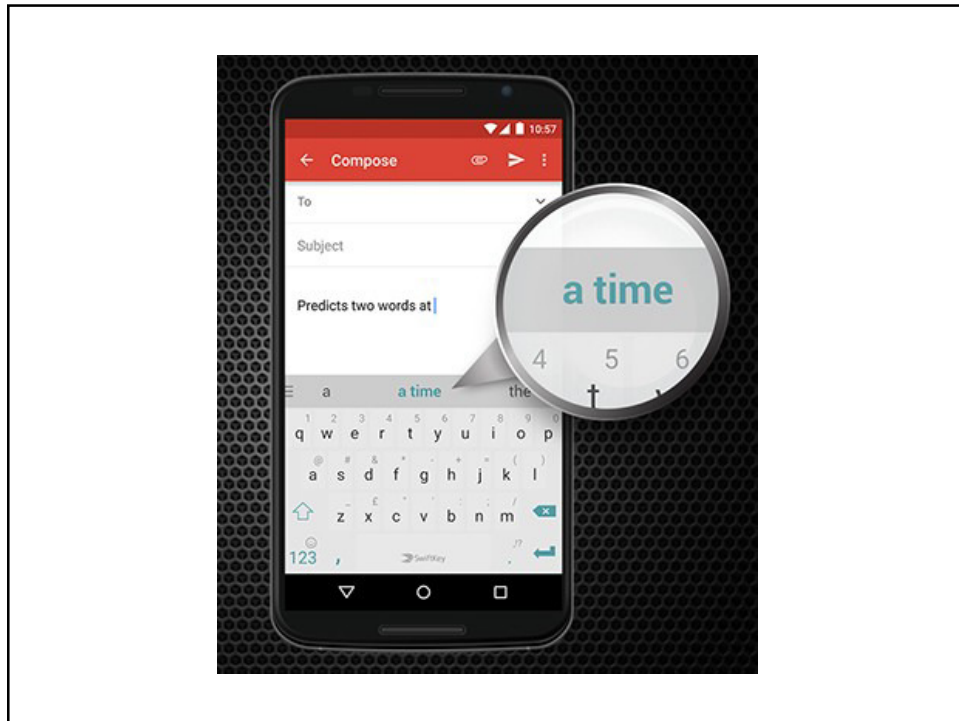
Datateknik med inriktning mot språkteknologi på KTH CSC
www.csc.kth.se/utbildning/kth/kompetensinriktning/sprakteknologi/
 Datateknik med inriktning mot **språkteknologi**. - det bästa av två världar. Finns det en gryende humanist i dig så är det här inriktningen för dig. Har du kanske ...

KTH | Språkteknologi (DD2418)
www.kth.se/Kurswebb
 Kursen ger en introduktion till **språkteknologi**. Du får lära dig: förklara och använda begrepp inom språkvetenskapens grundläggande nivåer: morfologi, syntax, ...
 Du har besökt den här sidan 2 gånger. Sidan besöktes senast: 2014-10-20

Utbildning i språkteknologi på KTH CSC
www.nada.kth.se/utbildning/sprakteknologi/







Watson (2011)



Vad behöver Watson kunna?

- Identifiera orden i kontinuerligt, fritt tal
- Analysera innebörden av ord och meningar
- Förstå flertydigheter och humoristiska formuleringar
- Blixtnabbt leta i en stor kunskapsbank för att hitta möjliga svar
- Välja ut rätt svar bland en mängd kandidater
- Läsa upp svaret med syntetiskt tal
- (I vissa situationer) Avgöra hur mycket pengar som ska satsas

Varför är språkteknologi svårt?

- Naturliga språk har flera egenskaper som formella språk (som Java, UML, predikatlogik) saknar:
 - Flertydighet
 - Vaghet
 - Ofullständighet
 - Dolda meningar, undertoner
 - Humor, ironi, ...
 - Ingen klar definition av vad som tillhör och inte tillhör språket
 - Naturliga språk utvecklas (spontant)

Lexikal flertydighet

- Ett ord kan som bekant ha flera betydelser:
 - Jag föredrar slips framför fluga.
 - Det satt en fluga på väggen.
 - Internet är bara en fluga.
- Ca 60% av svenskans ord har flera betydelser.

Lexikal flertydighet

- Vilken ordklass tillhör ordet "så"?
 - Så gick det till. (adverb)
 - På så sätt! (pronomen)
 - Han åt så han blev mätt. (konjunktion)
 - Så, det var intressant att höra! (interjektion)
 - Man måste så innan man kan skörda. (verb)
 - Grisarna drack ur en så. (substantiv)
- Att bestämma ordklass kan vara mycket svårt:
 - Han körde på vägen. (preposition)
 - Han körde på väggen. (partikel)

Lexikal flertydighet

- Hur många tolkningar har följande mening:
 - **"Time flies like an arrow"**

Strukturell flertydighet

- Fraser:
 - Filipa och Fredrik Reinfeldt
 - Silvia och Fredrik Reinfeldt
- Vart hör prepositionsfrasen?
 - Han såg flickan med det röda håret.
 - Han såg stjärnan med blotta ögat.
 - Han såg flickan med kikaren.
 - Erik gjorde pajen i köket.
 - Erik gjorde pajen i kylskåpet.

Strukturell flertydighet

I once shot an elephant in my pajamas. How he got into my pajamas I will never know”.

—Groucho Marx

Semantisk flertydighet

- Vem träffade Henrik?
 - Henrik träffade Eva.
 - Eva träffade Henrik. (topikalisering)
- Hur många personer?
 - Hon ville att hennes son skulle lära sig franska.
- Vad gick sönder?
 - Han tappade fatet på golvet och gjorde sönder det.
 - Han tappade stenen på tallriken och gjorde sönder den.

Pragmatisk flertydighet

- Kan du skicka saltet?
- Kan du hjälpa mig med det här?
- Det var ju jättefint det här.
- Duktig dator!

Metodologi

- **Handskrivna regler:**
 - Bygger på vår intuition om språket
 - Noam Chomsky: *Syntactic structures*, 1957.
- **Statistiska/datadrivna approcher:**
 - Tidigare approach än den regelbaserade
 - Har fått ett stort uppsving på senare år i samband med utveckling av hårdvara och mjukvara
- **Hybrid-approcher**
 - kombinerar kunskap om världen, kunskap om språk och statistik