

Sprachverarbeitung in der Informatik

Eine ganz kurze Einführung

Dieter Merkl

e-Commerce Arbeitsgruppe
Institut für Softwaretechnik und Interaktive Systeme
Technische Universität Wien
Favoritenstraße 9-11/188, A-1040 Wien

<http://www.ec.tuwien.ac.at>



Das Ziel

- Durch Einsatz von Künstlicher Intelligenz soll die Technik menschenfreundlicher gemacht werden
- Der Mensch soll sich nicht dem Computer anpassen müssen, sondern umgekehrt soll sich der Computer dem Menschen anpassen.

Sprache

- Für den Menschen schwer zu erzeugen, für den Computer leicht zu verstehen - Künstliche Sprache (z.B. C, Java, ...)

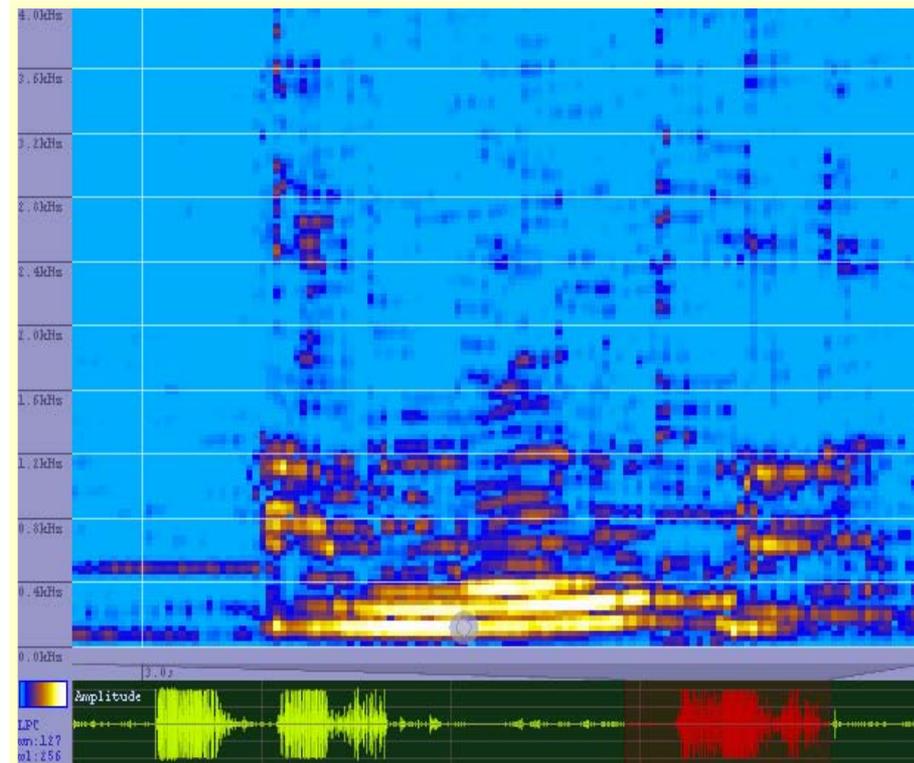
```
void norm_vect(double *vector, int dim) {  
    int i; double length=0.0;  
    for(i=0; i<dim; i++) { length += pow(vector[i], 2.0); }  
    if(length>0.0) { length = (double) 1/sqrt(length); }  
    for(i=0; i<dim; i++) { vector[i] *= length; }  
}
```

- Für den Menschen leicht zu erzeugen, für den Computer schwer zu verstehen - Natürliche Sprache (z.B. Deutsch, Tschechisch, ...)

Vítejte na přednášce! Jak se máš?

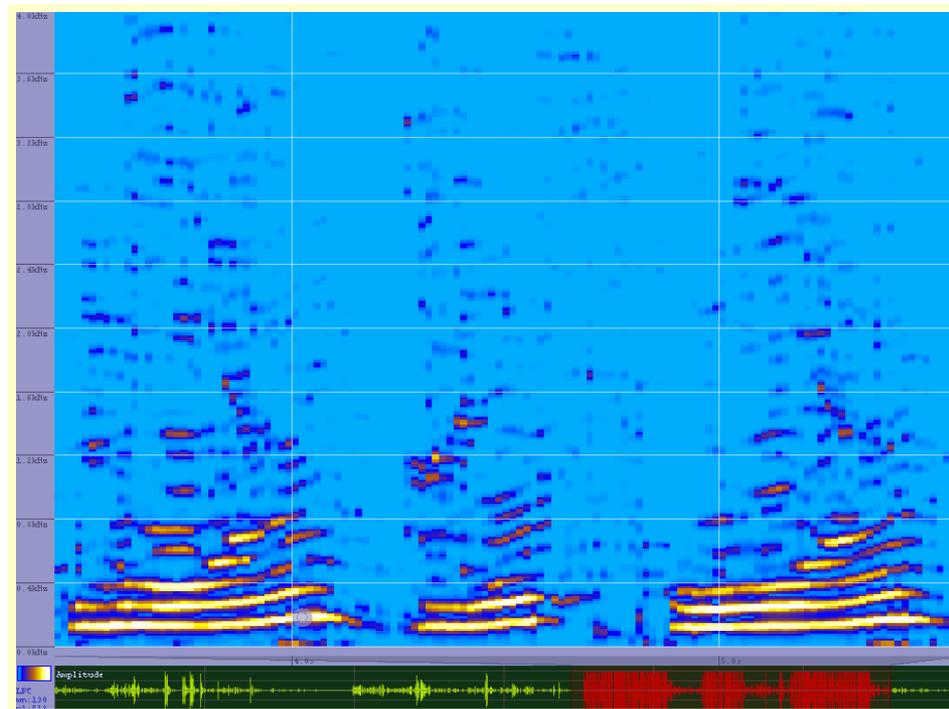
Warum ist Sprachverstehen für Computer so schwer?

- Wortgrenzen gehen im Sprachfluß unter
"amontag" -> "am Montag"



Warum ist Sprachverstehen für Computer so schwer?

- Gleiche Schallwellen werden je nach Kontext zu verschiedenen Wörtern
“Urlauber wollen wieder mehr ans Meer”



Warum ist Sprachverstehen für Computer so schwer?

- Der Mensch spricht "ohne Punkt und Komma"
"So machen wir das vielleicht klappt es"
 - > "So machen wir das. Vielleicht klappt es."
 - > "So machen wir das vielleicht. Klappt es?"
- Viele Menschen sprechen Dialekt
"I find des nätt"
 - > "Ich finde das nett"
 - > "Ich finde das nicht"
- Bei spontaner Rede entstehen viele Versprecher
"Wir treffen uns dann am Mon, äh, am Dienstag"

Warum ist Sprachverstehen für Computer so schwer?

- Dialogpartner fallen einander oft ins Wort
“Können wir uns am Mittwoch zusammen zum Essen ...”
“Da kann ich nicht!”
- Viele Formulierungen sind mehrdeutig
“Wir telefonieren mit Freunden in Prag”
 - > Wir telefonieren (mit Freunden in Prag)
 - > (Wir telefonieren mit Freunden) in Prag

INFORMATIK

Künstliche Intelligenz

Wie können die für das Sprachverstehen benötigten kognitiven Leistungen maschinell verfügbar gemacht werden?

Psycholinguistik
Wie wird Sprache von Menschen verstanden und generiert?

Sprachver- arbeitung

Linguistik
Wie können Sprachstrukturen mathematisch-logisch beschrieben werden?

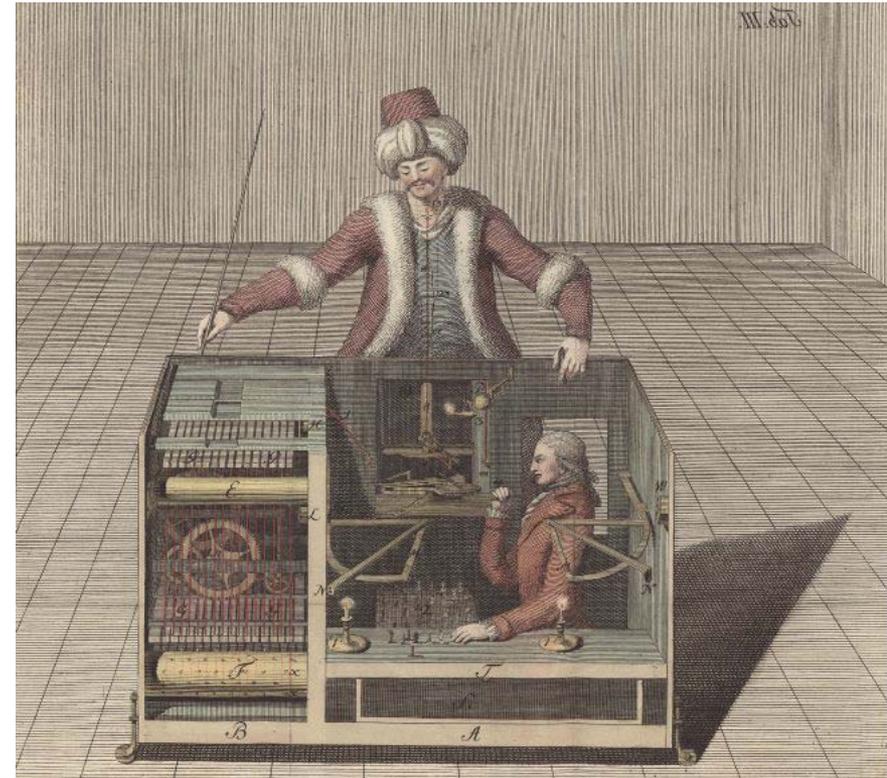
**KOGNITIONS-
WISSENSCHAFT**

Sprachtechnologie
Wie können effiziente, sprachverarbeitende und generierende Softwaresysteme erstellt werden?

Johann Wolfgang Ritter von Kempelen de Pázmánd



- * 23. Januar 1734 in Pressburg
† 26. März 1804 in der Alservorstadt, heute Wien
- Der Schachtürke (~ 1770)



Johann Wolfgang Ritter von Kempelen de Pázmánd

- Die Sprechmaschine
- *Mechanismus der menschlichen Sprache nebst der Beschreibung einer sprechenden Maschine (1791)*



Johann Wolfgang Ritter von Kempelen de Pázmánd

- Die Funktionsweise basierte auf einem möglichst naturgetreuen Nachbau der menschlichen Sprechorgane. Dabei wird die Lunge durch einen Blasebalg, die Funktion der Stimmbänder durch ein aufschlagendes Rohrblatt aus Elfenbein, die Nase durch zwei zusätzliche kleine Rohre und der Mund durch einen Gummitrichter simuliert. Durch Veränderung der Abdeckung des Gummitrichters mit der Hand lassen sich sowohl unterschiedliche Vokale als auch gewisse Konsonanten erzeugen. Mangels ausreichendem Resonanzraumes konnten nicht alle Sprechlaute nachgebildet werden.
- Diesen Mangel verdeckte Kempelen dadurch, dass er seiner Sprechmaschine eine Kinderstimme gab, die den Zuhörer in Bereitwilligkeit versetzt, eine mangelhafte Artikulation zu tolerieren.

Joseph Weizenbaum



- * 8. Januar 1923 in Berlin
- † 5. März 2008 in Gröben bei Berlin

- ELIZA (1966)
- ELIZA wurde als Meilenstein der „künstlichen Intelligenz“ gefeiert
- Weizenbaum war entsetzt, wie ernst viele Menschen dieses relativ einfache Programm nahmen
- Durch dieses Schlüsselerlebnis wurde Weizenbaum zum Kritiker der gedankenlosen Computergläubigkeit
- *Die Macht der Computer und die Ohnmacht der Vernunft*, Suhrkamp, Frankfurt am Main, 1977

Joseph Weizenbaum

- ELIZA: Name in Anlehnung an das Schauspiel Pygmalion von George Bernard Shaw
- Prinzip: Aussagen des menschlichen Gesprächspartners in Fragen umzuformulieren
 - Benutzer: „Ich habe ein Problem mit meinem Auto.“
 - ELIZA: „Warum, sagen Sie, haben Sie ein Problem mit Ihrem Auto?“
- Erkennen von Schlüsselwörtern und Reaktion darauf
 - Benutzer: „Ich habe ein Problem mit meinem Vater.“
 - ELIZA: „Erzählen Sie mir mehr über Ihre Familie!“
- Soweit so gut
 - Benutzer: „Krieg ist der Vater aller Dinge.“
 - ELIZA: „Erzählen Sie mir mehr über Ihre Familie!“

Terry Winograd



- * 24. Februar 1946 in Takoma Park, MD
- <http://hci.stanford.edu/~winograd/>
- SHRDLU (1970)
- PhD Thesis am MIT
- Erstellen eines Computerprogrammes, das über ausreichend viel "Weltwissen" verfügt, dass man damit in natürlicher Sprache interagieren kann
- Die Welt von SHRDLU: Bauklötze :-)



Terry Winograd

- Person: Pick up a big red block.
- Computer: OK.
- Person: Grasp the pyramid.
- Computer: I don't understand which pyramid you mean.
- Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
- Computer: By "it", I assume you mean the block which is taller than the one I am holding.
- Computer: OK.
- Person: What does the box contain?
- Computer: The blue pyramid and the blue block.
- Person: What is the pyramid supported by?
- Computer: The box.
- Person: How many blocks are not in the box?
- Computer: Four of them.

Text Retrieval

- Focus auf Textsuche
- Statistische Methoden zur Textanalyse
- Suchmaschinen speichern Information über Web-Seiten
- Web-Seiten müssen gefunden und geladen werden
 - > Crawler, Spider
- Web-Seiten müssen analysiert werden, d.h. eine Beschreibung des Inhalts wird erzeugt
 - > Indexer
- Ausgehend von Suchanfragen werden passende Web-Seiten als Suchergebnis präsentiert
 - > Search-Interface

who was the first man on the moon - Google Search

http://www.google.com/search?hl=en&q=who-☆

Most Visited Latest Headlines TWiki Apollo Amazon Standard WebMail ECTWiki LEO tuwel TUWIS

Web Images Maps News Shopping Gmail more Sign in

Google who was the first man on the moon Search Advanced Search Preferences

Web Results 1 - 10 of about 2,710,000 for who was the first man on the moon. (0.19 seconds)

BBC ON THIS DAY | 21 | 1969: Man takes first steps on the Moon
American Neil Armstrong becomes the **first man** to walk on the **Moon**.
news.bbc.co.uk/onthisday/hi/dates/stories/july/21/newsid_2635000/2635845.stm - 33k - [Cached](#) - [Similar pages](#)

Apollo 11 First man on the Moon-First Moon Landing Neil Armstrong ...
Panorama with audio from the **first moon** landing Apollo 11.
www.panoramas.dk/moon/apollo-11.html - 22k - [Cached](#) - [Similar pages](#)

Apollo 11-17 First man on the Moon - Apollo Moon Landings - QTVR ...
The Apollo Misions - Nasa in Space - Experience the **moon** just as the astronauts did.
www.panoramas.dk/FULLSCREEN3/F29.HTML - 57k - [Cached](#) - [Similar pages](#)

First Man On The Moon
First Man On The Moon. Who filmed Neil Armstrong coming out of the Lunar Module when there was no-one on the surface to do this? ...
www.redzero.demon.co.uk/moonhoax/First_Man.htm - 6k - [Cached](#) - [Similar pages](#)

Apollo 11 - Wikipedia, the free encyclopedia
Chaikin, Andrew (1998). A **Man on the Moon**. Penguin Group, pp. 204 & 623. ... **First moon** landing in 1969 marked an entire generation - Matt Wallace - Opinion ...
en.wikipedia.org/wiki/Apollo_11 - 123k - [Cached](#) - [Similar pages](#)

Being The First Man On The Moon, Ed Bradley Talks To Neil ...
In the **first** television profile he has ever agreed to, Neil Armstrong tells Ed Bradley he doesn't deserve the celebrity that comes with being the **first man** ...
www.cbsnews.com/stories/2005/11/03/60minutes/main1008288.shtml - 98k - [Cached](#) - [Similar pages](#)

Hollywood hunts star to play first man on the Moon | Film | The ...
Producers search for a talent big enough to fill Neil Armstrong's boots in film of astronaut's life.
www.guardian.co.uk/film/2008/apr/13/news.spaceexploration - 80k - [Cached](#) - [Similar pages](#)

Sponsored Links
Neil Armstrong Museum
Learn about the **first moon** landing & the NASA programs that led to it!
www.ohiohistory.org

Done

Neue Entwicklungen

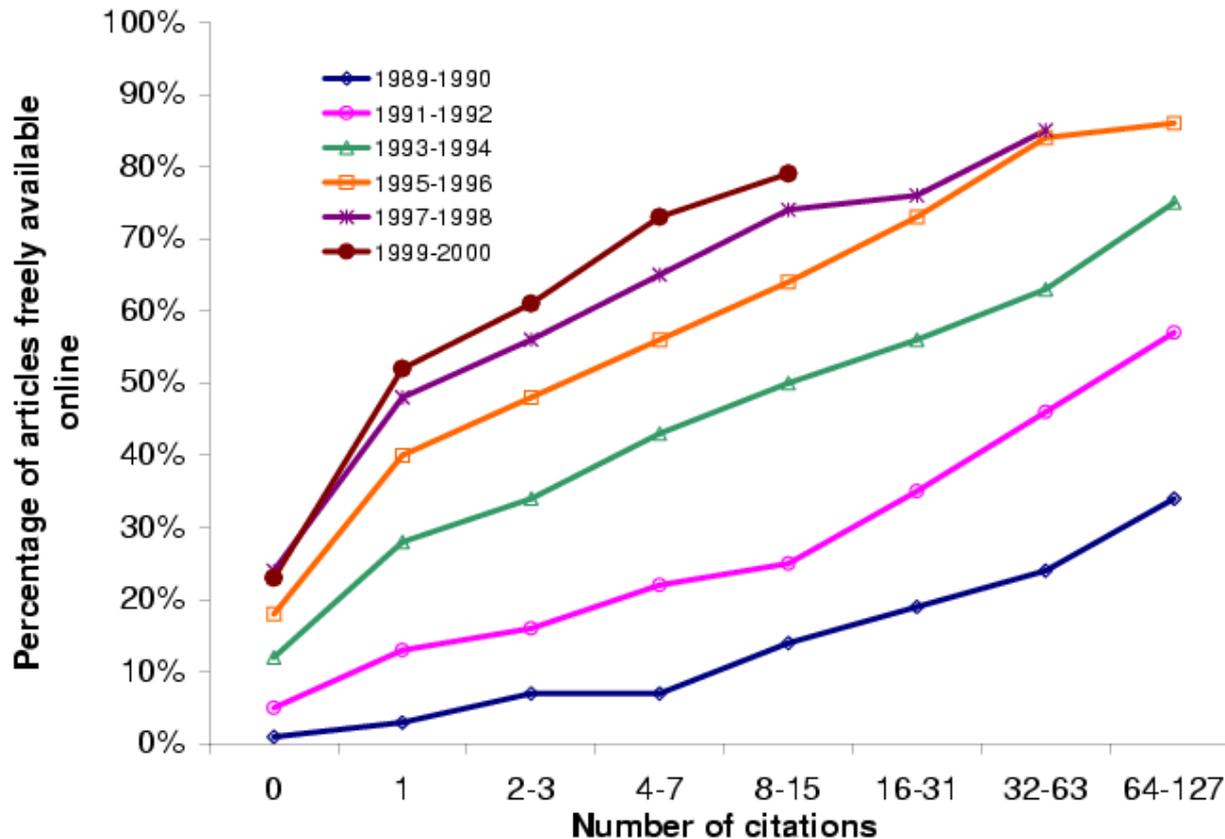
- Spezialisierte Suchmaschinen
- “Antwortmaschinen” statt “Suchmaschinen”
- Übersetzungssysteme

CiteSeer

- 1997 NEC Research Institute, Princeton, NJ
- Suchmaschine für wissenschaftliche Literatur in der Informatik
- Grundannahmen
 - Autoren wissenschaftlicher Arbeiten hätten es gerne, wenn ihre Arbeiten gelesen werden
 - Eine wissenschaftliche Arbeit folgt einem bestimmten Aufbau
 - Über Referenzen lassen sich weitere wissenschaftliche Arbeiten finden
- Und der Rest ist ein einfacher Schneeball Effekt :-)

CiteSeer

- Steve Lawrence: Online or Invisible, Nature, Vol 411, No 6837, 2001, p 521.



Lexxe

- 2005 in Sydney gegründet
- Verwendet Sprachanalysemethoden, um Antworten auf Anfragen in Web-Dokumenten zu finden
- Ergebnisdokumente werden zu thematischen Gruppen zusammengefasst

Lexxe

- what is the name of the capital of australia

The screenshot shows a web browser window with the address bar containing the URL: `http://www.lexxe.com/main.cfm?sstring=what+is+the+name+of+the+capital+of+australia&clickcluster=fmclk`. The search engine's logo is "LexxeTMalpha" with the tagline "powered by advanced natural language technology". The search input field contains the query "what is the name of the capital of australia" and the "Find Answer" button is highlighted. A blue arrow points from the "Find Answer" button to the search result. The search results are displayed in a table with columns "Cluster:" and "Web:". The first result is "Canberra - Wikipedia, the free encyclopedia" with a brief description and coordinates. Other results include "Canberra National", "Canberra Australia", and "Sydney and Melbourne". A sidebar on the right contains promotional text for Lexxe.

Lexxe Search Engine: [Q:] what is the name of the capital of australia [A:] Canberra

http://www.lexxe.com/main.cfm?sstring=what+is+the+name+of+the+capital+of+australia&clickcluster=fmclk

Headlines TWiki Standard LEO CiteSeer Amazon Gmail Apollo Stadtplan EeePC Wiki

LexxeTMalpha
powered by advanced natural language technology

Please type your question or key words below >
what is the name of the capital of australia

Find Answer Help

Tip: You can simply use the "define" command to obtain a definition of a word or phrase, e.g. define: hygiene.

Answer: Canberra

Cluster: Web: 100 results found. Definitions: name capital au

- > Canberra Capital of Australia
Canberra - Wikipedia, the free encyclopedia
For other uses, see Canberra (disambiguation). Canberra. Australian Capital ...
35°18' 27' ' S 149°07' 27.9' ' E / -35.3075, 149.124417 Coordinates: 35°18' 27' ' S 149° ...
<http://en.wikipedia.org/wiki/Canberra> - 160 KB
- > Canberra National
- > Canberra Australia
Canberra Australia - Find Your Way to Parliament House
Canberra, capital of Australia, is a compact planned Australian city conducive to short-distance sightseeing. ... Discover Australia Discover New Zealand ...
<http://goaustralia.about.com/cs/actsightseeing/a/canberra1.htm> - 23 KB
- > Sydney and Melbourne

Do you like Lexxe?
If you think it's cool and we help you, please help Lexxe, please tell your friends about Lexxe, please email us and blog to www.lexxe.com.

Invest in Lexxe
We really like Lexxe.

Done

Lexxe

- what is 42

The screenshot shows a web browser window with the Lexxe Search Engine interface. The search query is "what is 42" and the answer is "<1> being two more than forty;". The search results are displayed in a table with columns for "Cluster:" and "Web:". The first result is "Guitar That Delivers Everything You'd Expect" from Wikipedia, titled "Answer to Life, the Universe, and Everything". The second result is "Class Lcd Hdtv" from Wikipedia, titled "42 (number)". The third result is "Shop Find Buy" from Wikipedia, titled "The New General Catalogue object NGC 42".

Lexxe Search Engine: [Q:] what is 42 [A:] <1> being two more than forty;

http://www.lexxe.com/main.cfm?sstring=what+is+42&clickcluster=fmclk&sstringtemp=fmstr

Headlines TWiki Standard LEO CiteSeer Amazon Gmail Apollo Stadtplan EeePC Wiki

LexxeTM**alpha**
powered by advanced natural language technology

Please type your question or key words below >
what is 42

Find Answer Help

Tip: You can simply use the "define" command to obtain a definition of a word or phrase, e.g. define: hygiene.

Answer: <1> being two more than forty;

Cluster: Web: 100 results found. Def

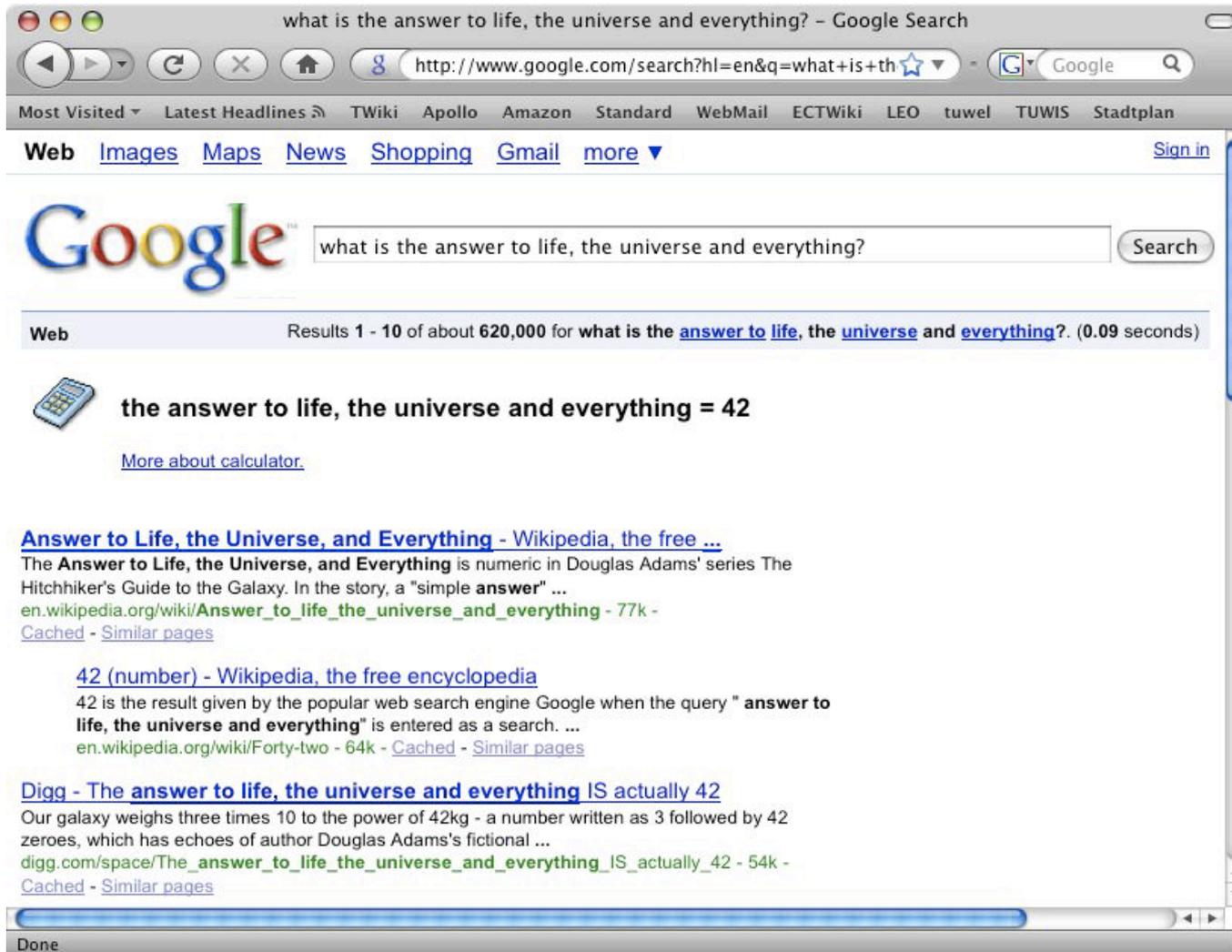
- > **Guitar That Delivers Everything You'd Expect**
Answer to Life, the Universe, and Everything - Wikipedia, the free ...
Prak confirms that 42 is indeed The Ultimate Answer, and confirms that it is ... for The Burkiss Way called "42 Logical Positivism Avenue", broadcast on BBC ...
http://en.wikipedia.org/wiki/The_Answer_to_Life,_the_Universe,_and_Everything - 83 KB
- > **Class Lcd Hdtv**
42 (number) - Wikipedia, the free encyclopedia
The New General Catalogue object NGC 42, a spiral galaxy in the constellation Pegasus ... for the author Douglas Adams who popularized the number 42 and died in 2001. ...
[http://en.wikipedia.org/wiki/42_\(number\)](http://en.wikipedia.org/wiki/42_(number)) - 55 KB
- > **Shop Find Buy**
Plasma Tv

Do you like Lexxe?
If you think cool and v help
Lexxe, please your friends about email and blog to www.lexxe.com

Invest in We really

Done

PS: bei Google übrigens auch ;-)



Google Language Tools



Google Language Tools

- Dieter Merkl is Associate Professor at Vienna University of Technology. His current interests are in the areas of data mining, art history, information retrieval, architecture, neural computation and French cinema. In some of these areas, Dieter has published more than 100 papers in conference proceedings and journals. Dieter received his master's and doctoral degrees in Social and Economic Sciences from the University of Vienna.
- Dieter Merkl studierte Betriebs- und Wirtschaftsinformatik an der Universität Wien. Nach seinem Studium war er zunächst Projektmitarbeiter, später Universitätsassistent am Institut für Informationssysteme der Universität Wien. Seit 1995 ist er am Institut für Softwaretechnik und Interaktive Systeme beschäftigt. Er hat sich 1998 an der Technischen Universität Wien für das Fach Angewandte Informatik habilitiert mit Arbeiten im Bereich des Text Mining, insbesondere Dokumenten-Clustering. In den Jahren 2000 - 2003 war er Projektleiter für Adaptive Multilinguale Interfaces am K-ind Electronic Commerce Competence Center (EC3). In den Jahren 2004 - 2005 war Dieter Merkl als Associate Professor for Software and Systems Engineering an der University of Western Sydney tätig. Aktuelle Forschungsschwerpunkte von Dieter Merkl sind Kooperative Virtuelle Umgebungen, Interaction Design, Data Mining und e-Commerce. In diesen Bereichen hat er mehr als 140 Arbeiten in wissenschaftlichen Zeitschriften bzw. bei wissenschaftlichen Tagungen veröffentlicht. Eines seiner aktuellen Forschungsprojekte ist das FWF-Translational Research Projekt 3D e-Tourismus Umgebung.

Google Language Tools

Original

Dieter Merkl is Associate Professor at Vienna University of Technology. His current interests are in the areas of data mining, art history, information retrieval, architecture, neural computation and French cinema. In some of these areas, Dieter has published more than 100 papers in conference proceedings and journals. Dieter received his master's and doctoral degrees in Social and Economic Sciences from the University of Vienna.

english2german

Dieter Merkl ist Associate Professor an der Technischen Universität Wien. Seine aktuellen Interessen liegen in den Bereichen Data Mining, Kunstgeschichte, Information Retrieval, Architektur, Neural Computation und französischen Kinos. In einigen dieser Bereiche, Dieter hat mehr als 100 Zeitungen in Tagungsband und Zeitschriften. Dieter erhielt seinen Master-und Doktoranden-Grad in Sozial-und Wirtschaftswissenschaften an der Universität Wien.

english2french

Dieter Merkl est professeur associé à l'Université de technologie de Vienne. Ses intérêts actuels sont dans les domaines de l'extraction de données, histoire de l'art, la recherche d'information, de l'architecture, de neurones et de calcul du cinéma français. Dans certains de ces domaines, Dieter a publié plus de 100 articles dans actes de congrès et de revues. Dieter a reçu sa maîtrise et son doctorat en Sciences économiques et sociales de l'Université de Vienne.

english2czech

Dieter Merkl je docent na Vienna University of Technology. Jeho zájmy jsou aktuální v oblasti data miningu, dějiny umění, vyhledávání informací, architektura, neuronové výpočty a francouzské kinematografie. V některých z těchto oblastí, Dieter vydala více než 100 dokumentů v konferenčních sbornících a časopisech. Dieter přijali jeho diplomových a doktorských studijních programů pro sociální a ekonomické vědy na univerzitě ve Vídni.

Google Language Tools

Original

Dieter Merkl studierte Betriebs- und Wirtschaftsinformatik an der Universität Wien. Nach seinem Studium war er zunächst Projektmitarbeiter, später Universitätsassistent am Institut für Informationssysteme der Universität Wien. Seit 1995 ist er am Institut für Softwaretechnik und Interaktive Systeme beschäftigt. Er hat sich 1998 an der Technischen Universität Wien für das Fach Angewandte Informatik habilitiert mit Arbeiten im Bereich des Text Mining, insbesondere Dokumenten-Clustering. In den Jahren 2000 - 2003 war er Projektleiter für Adaptive Multilinguale Interfaces am K-ind Electronic Commerce Competence Center (EC3). In den Jahren 2004 - 2005 war Dieter Merkl als Associate Professor for Software and Systems Engineering an der University of Western Sydney tätig. Aktuelle Forschungsschwerpunkte von Dieter Merkl sind Kooperative Virtuelle Umgebungen, Interaction Design, Data Mining und e-Commerce. In diesen Bereichen hat er mehr als 140 Arbeiten in wissenschaftlichen Zeitschriften bzw. bei wissenschaftlichen Tagungen veröffentlicht. Eines seiner aktuellen Forschungsprojekte ist das FWF-Translational Research Projekt 3D e-Tourismus Umgebung.

german2english

Dieter Merkl studied business and marketing computer science at the University of Vienna. After his studies he was initially project staff, and later Assistant Professor at the Institute for Information Systems at the University of Vienna. Since 1995 he has at the Institute for Software Technology and Interactive Systems employs. He has in 1998 at the Technical University of Vienna for applied computer science specialist with habilitation work in the field of text mining, clustering in particular documents. In the years 2000 - 2003 he was project manager for Adaptive Multilingual interfaces at the K-ind Electronic Commerce Competence Center (EC3). In the years 2004 - 2005 was Dieter Merkl as an Associate Professor for Software and Systems Engineering at the University of Western Sydney operates. Current research priorities by Dieter Merkl are Collaborative Virtual Environments, Interaction Design, data mining and e-commerce. In these areas, he has more than 140 works in scientific journals and at scientific meetings published. One of his current projects is the FWF Translational Research Project 3D e-tourism environment.

Google Language Tools

german2french

Dieter Merkl des études d'exploitation et de l'informatique de l'Université de Vienne. Après ses études, il a été d'abord collaborateur de projet, plus tard, Assistant à l'Institut de systèmes d'information de l'Université de Vienne. Depuis 1995, il est à l'Institut de technologie de la programmation interactive et de systèmes de traite. Il a, en 1998 à l'Université technique de Vienne, les cours d'informatique appliquée thèse d'habilitation à travailler dans le domaine du Text Mining, et notamment les documents de clustering. Dans les années 2000 - 2003, il a été chef de projet pour Adaptive interface multilingue, le K-ind Electronic Commerce Competence Center (EC3). Dans les années 2004 - 2005 a été Dieter Merkl professeur associé for Software and Systems Engineering à l'University of Western Sydney activité. Priorités de recherche, en cours de Dieter Merkl coopérative sont des environnements virtuels, Interaction Design, Data Mining et e-commerce. Dans ces domaines, il a plus de 140 travaux dans des revues scientifiques ou de réunions scientifiques publiés. L'un de ses projets de recherche en cours est le FWF-projet de recherche translationnelle 3D e-tourisme environnement.

german2czech

Dieter Merkl studoval obchodní a marketingové počítačové vědy na univerzitě ve Vídni. Po studiích byl původně projekt zaměstnanců, a později odborný asistent na Ústavu informačních systémů na univerzitě ve Vídni. Od roku 1995 byl v ústavu pro Softwarové technologie a interaktivní systémy zaměstnává. Byl v roce 1998 na Technické univerzitě ve Vídni pro aplikovaný počítačové vědy specialista s habilitační práci v oboru text hornické činnosti, zejména shlukování dokumentů. V letech 2000 - 2003 byl jako projektový manažer pro adaptivní rozhraní Multilingual na K-ind Electronic Commerce Competence Center (EC3). V letech 2004 - 2005 byl Dieter Merkl jako docent na software a Systémové inženýrství na Západočeské univerzitě v Sydney funguje. Aktuální výzkumné priority, které jsou Dieter Merkl Kolaborativní virtuální prostředí, interakce Design, dolování dat a e-komerce. V těchto oblastech, má více než 140 prací ve vědeckých časopisech a na vědeckých konferencích zveřejněny. Jedním z jeho aktuálních projektů je FWF Translační výzkum Projekt 3D e-cestovní ruch prostředí.

Gibt's Fragen?



Q&A QUESTIONS & ANSWERS

May 24-25 Questions

- ☁ Why doesn't God answer my prayers?
- ☁ How does Southeast plan to punish Shawn Wells for using 3 parking spaces on Sunday Mornings?
- ☁ The Trinity; I understand the concept, but could you go a little more in-depth?
- ☁ Why do boys have nipples?
- ☁ What does the Bible say about homosexuality?
- ☁ Will you go to prom with me?



Quellen

- W. Wahlster: Perspektiven in der Sprachtechnologie: Hörende, sprechende und verstehende Computer
www.st.cs.uni-sb.de/edu/perspektiven/wahlster.ppt
- Informationen zu Johann Wolfgang Ritter von Kempelen de Pázmánd, Joseph Weizenbaum, ELIZA, Terry Winograd
siehe Wikipedia (de.wikipedia.org)



FAKULTÄT FÜR **INFORMATIK**

ec *electronic commerce group*

ao.Univ.Prof. Dr. Dieter Merkl

e-Commerce Arbeitsgruppe

Institut für Softwaretechnik und Interaktive Systeme

Technische Universität Wien

Favoritenstraße 9-11/188, A-1040 Wien

dieter.merkl@ec.tuwien.ac.at

www.ec.tuwien.ac.at/~dieter