

Compass: An Information Retrieval Tool

by

Yannick Assogba

In fulfillment of the requirements of

COMP 490

Concordia university

May 2006

Introduction

This project introduces Compass, Compass is an automatic information retrieval tool, designed to enhance information retrieval from the web, particularly by improving the results returned by common web search engines such as Google or Yahoo. The approach it takes considers the observation that human search queries are often not generated, so to speak, in a vacuum; by this I mean that search queries are often generated in a larger context relating to the users information retrieval needs. The context that Compass is designed to work with is that of the document. Searches in Compass are done directly from a source document; the source document provides the context representing the user's information retrieval goals and this context is used to tailor search results to the perceived user needs.

The figure below (figure 1) depicts a typical result page of a popular web search engine, in this case Yahoo. The search terms used were "moles" and "family", the intent being to find out more about moles (the small furry mammal) and the members of their family (Talpidae). However not all the results reflect my information needs, and the topics referenced range from the Moles family, to spies and funeral homes.

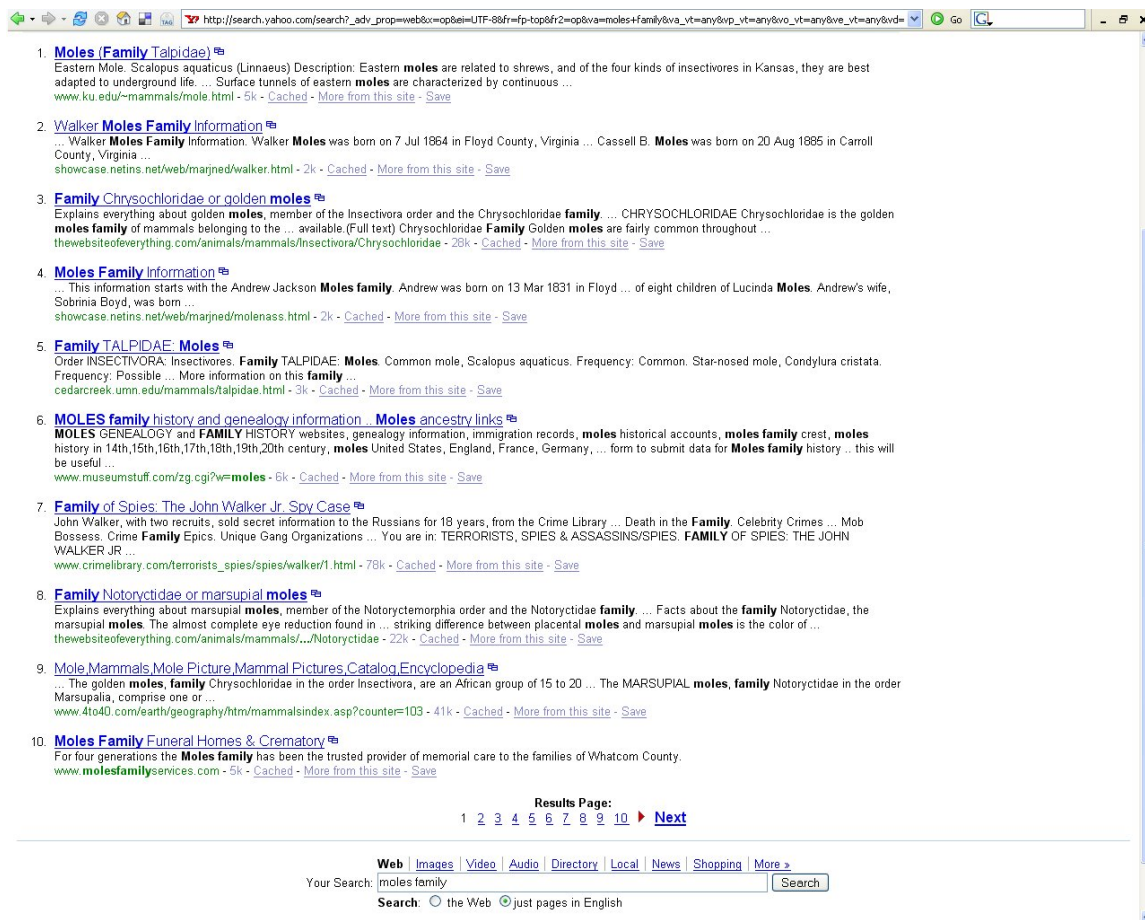


Figure 1: Yahoo search results

Ideally the results returned should all be relevant to my goal, unfortunately Yahoo! doesn't know about my needs and the information provided (search terms) does not encode this appropriately. This is where Compass comes in, as mentioned before it uses documents to decode your search goals, helps you create search queries, and tailors the results to fit you goals. Figure 2 (below) shows the front end of compass; in it a document (about

moles) is open and the user has made a selection. From this selection the user can request Compass to “Get More Information”.

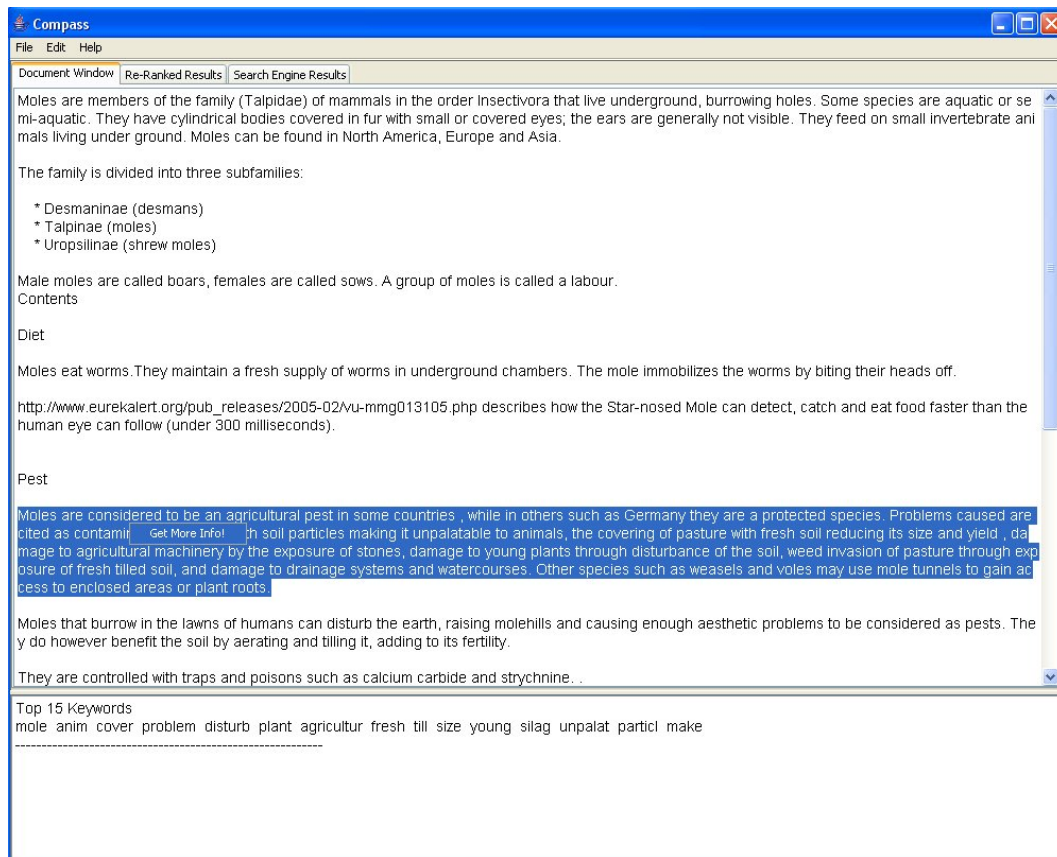


Figure 2: Compass front end

The process used in producing results can be described in four major steps:

- Document Analysis & Keyword Extraction
- Search Query Generation
- Web Search
- Result Document Ranking

This report aims to describe Compass' operation in detail and prefaces this with a discussion of the issues in information retrieval

that motivated this work. Finally, results and evaluation of Compass' performance are presented.

What is information retrieval?

Information Retrieval (IR) research encompasses the development of algorithms, techniques and models for the retrieval of information from document repositories. This definition would benefit further clarification as the term is wide and somewhat loosely defined; this report considers information retrieval from the perspective outlined by van Rijsbergen [5], he states *“An information retrieval system does not inform (i.e. change the knowledge of) the user on the subject of his inquiry. It merely informs on the existence (or non-existence) and whereabouts of documents relating to his request”*. Accordingly I will be discussing IR in the context of the retrieval of documents from document repositories, van Rijsbergen continues. *“In principle, information storage and retrieval is simple. Suppose there is a store of documents and a person (user of the store) formulates a question (request or query) to which the answer is a set of documents satisfying the information need expressed by his question. He can obtain the set by reading all the documents in the store, retaining the relevant documents and discarding all the others. In a sense, this constitutes 'perfect' retrieval. This solution is obviously impracticable. A user either does not have the time or does not wish to spend the time reading the entire document collection, apart from the fact that it may be physically impossible for him to do so.”*

The field of IR has received increasing attention since the 1940's as a tool in managing the explosion in scientific literature of the last few decades and IR systems are seen in an academic context quite frequently as most university libraries use them to provide access to their catalogue. For example Concordia University's library has an IR system called CLUES [CL] providing access to the library catalogue, the strategy used in CLUES and many other library IR systems allows a user to search through metadata describing the contents of the library, the metadata for each item in the catalogue is determined and entered into the system manually when that item is introduced into the library. A typical CLUES record allows the following information to be searched through: author, title, publisher, subject area and notes that can indicate the content of the item¹ (these are usually a set of keywords, manually extracted from the item). As long as the descriptors decided on previously for the item you are looking for match the descriptors you come up with at the time of your search, you should be able to find relevant information fairly painlessly. This may often not be the case and it seems quite easy to miss potentially relevant books and articles if they do not share the same metadata². Recently, the web has also become a viable repository for academic literature and the boon afforded by documents on the web, or indeed

1 Most CLUES records that the author looked at did not have a notes entry.

2 A search using the keywords "moles" and "family" returns no results. This may lead one to conclude that Concordia Library has no information on the family Talpidae.

by any document stored in an electronic medium, is that of full text search. Rather than only being able to search predetermined metadata, we can search the full contents of books, periodicals, articles and papers. This does help us greatly in the problem of missing documents completely within our result set. It however often creates a problem of returning too many results, often including some that are unrelated to our original search goals³.

3 This problem is also shared CLUES, a search using just the keyword “moles” returns 40 results, of which only 2 are about moles.

The ad-hoc retrieval problem

The classical problem in IR is the *ad-hoc retrieval problem*, in which the user enters a query describing the desired information and the system returns a list of documents, preferably in an order ranked from most to least relevant. Broadly speaking, the solution consists of at least two steps, the first being search and the second that of assigning relevance to documents with respect to a query i.e. result ranking.

When the document repository is large and heterogeneous, as is the case for the web, result sets of the search step are often huge and unwieldy. So while the web provides a rich and diverse data set from which to obtain information, the current tools available to users are less than perfect. Anyone that has spent considerable time doing research on the Internet using popular tools such as the Google and Yahoo! search engines have experienced the problem of search operations returning too many irrelevant documents. In IR terminology, this is known as the problem of *precision*, precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved [5] i.e.

$$\textit{precision} = (\textit{number of relevant documents retrieved}) / (\textit{number of documents retrieved})$$

A related measure is that of *recall*, recall measures the proportion of relevant documents returned with respect to the number of relevant documents contained within the repository. The description of CLUES given suggests that precision is high in this type of search, however, as noted, it intuitively seems quite possible to completely miss information on relevant material, i.e. recall is low. There is generally a trade-off between recall and precision [2] and web search typically swings in the opposite direction towards higher recall. Most commercial web search engines are based upon *exact match* retrieval systems. Exact match systems return documents that precisely satisfy some structured query of which the best known type is the Boolean query (used in Google and Yahoo! among others). Thus a typical search operation on these search engines will return all documents that contain at least one instance of *all* (or sometimes *any*) of the terms specified in the query, combined with full text search and a large repository it is evident that although recall in the result set may be high⁴, precision can often be low⁵.

Precision however does not capture another quality of a result set which is also desirable to the user. Those documents within the result

4 This is mostly because the results set are so large, they are likely to contain a lot of the relevant data within them somewhere. It should be noted that with the definition of recall given, if the result set contains the entire repository then recall is 100%.

5 The opposite situation is also possible, the result set is very small (or empty) indicating that recall is low while precision may likely improve.

set that are likely relevant to the users needs are only helpful if they are easily discoverable by the user. In IR this is encapsulated in the probability ranking principle which can be stated as follows, *“If a reference retrieval system's response to each request is a ranking of the documents in the collection in order of decreasing probability of relevance to the user who submitted the request, where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system for this purpose, the overall effectiveness of the system to its user will be the best that is obtainable on the basis of those data”* [5]. Thus the second step in ad-hoc retrieval aims to provide the most relevant results to the user as soon as possible. A poor ranking is just as frustrating as an imprecise search. It is this second problem, that of ranking documents in the result set, that this project addresses most directly.

Compass in operation

When a user opens a document (the source document) in Compass, it is analyzed and important words (*keywords – see below*) are extracted and stored, this corresponds to step 1 in the outline given previously.

The user is then free to select any portion of text which is indicative of their search needs. Once they have done this, they can right click and select the “get more information” option. Compass will use the keywords (extracted earlier) within the users selection to form a search query which is passed on to a web based search engine (steps 2 and 3 respectively). Compass currently works with Google and Yahoo! The result set is obtained from the search engine and the documents it points to are downloaded, analyzed and compared to the source document. Each document is assigned a score and the result set is re-ranked (step 4) to promote documents that are more closely related to the source document as a whole. This new result set is then presented to the user.

Keyword Extraction

Keywords are significant words that are descriptive of the text in which they occur. Significance here is related to how much a word describes the semantics of the document. This concept of keyword needs to be differentiated from that of search term, in traditional search engines search terms do not capture the notion of significance, for exact match systems all words in a document describe that document with equal weight. Thus keywords make for good search terms but not all search terms are 'keywords' with respect to the documents they retrieve (in the case where relatively insignificant words are used in search, e.g. “some”⁶). Consequently the appropriate selection of search terms is crucial to the precision of the result set in commercial search systems. If search terms are poorly selected then the result set will often be of low precision, on the other hand (due to the nature of exact match search) if too many specific words are used in a search query the size of the result set often decreases drastically and in some cases becomes empty.

The keyword extraction algorithm used in Compass is a partial implementation of an algorithm described in the paper *Keyword Extraction from a Single Document using Word Co-Occurrence*

⁶ The word “some” is not filtered out by Google's stop word remover

*Statistical Information by Matsuo & Ishizuka (2003) [3]. It is a statistics based algorithm that has the advantage of not requiring a corpus yet maintaining comparable performance to popular tf-idf (Term Frequency-Inverse Document Frequency) techniques that employ a corpus. These techniques work by capturing the frequency of occurrence of words in a representative corpus to which the frequency of word occurrence in the document is compared. Words that occur significantly more frequently in the document as compared to the corpus, are potential keywords. These techniques suffer from the disadvantage that they require a corpus and are language dependent; in fact they are often tied to the domain from which the corpus originated. The algorithm used in Compass does not suffer from these disadvantages⁷. More information on tf-idf techniques can be found in [2]. As my implementation is only partial I cannot claim the same performance for my keyword extraction, however empirical data shows my implementation to be effective for the task. Matsuo & Ishizuka describe the approach as follows. “Frequent terms are extracted first, then a set of co-occurrences between each term and the frequent terms, i.e., occurrences in the same sentences, is generated. Co-occurrence distribution shows importance of a term in the document as follows. If the probability distribution of co-occurrence between term **a** and the frequent terms is biased to a*

⁷ The stop word list is language dependent but it can be changed extremely easily with no training required.

*particular subset of frequent terms, then term **a** is likely to be a keyword.” [3]*

In brief if a word co-occurs selectively with certain frequent terms it is more likely to be a significant word than a word that co-occurs fairly evenly with respect to the frequent terms. The χ^2 measure is used as a measure of co-occurrence bias. The algorithm as implemented by the author is described in greater detail below.

Algorithm

- 1. Preprocessing:** Words are stemmed using the Porter stemming algorithm [4], the full forms of the words are stored along with their stem but it is the stemmed version that is used in further statistical calculations. Discard words occurring in the stop list [see appendix A].
- 2. Selection of frequent terms:** Select the top frequently occurring terms, up to 30% of the total number of unique terms.
- 3. Calculation of χ^2 value:** The χ^2 value is the measure of bias used to score words, words with a higher χ^2 value (i.e. words that are more biased toward a particular subset of the frequent terms), are more likely to be keywords. The χ^2 value of a term w is given by the following equation.

$$\chi^2(w) = \sum_{g \in G} \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g} - \max_{g \in G} \left\{ \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g} \right\}$$

Where: G = set of frequent terms

$\text{freq}(w, g)$ = co-occurrence frequency of a term w with a frequent term g

n_w = total number of terms in sentences including w

$\mathbf{p_g}$ = sum of total number of terms in sentences in which
g occurs divided by the total number of terms

4. Return Keywords: Terms are ordered by their x^2 value and the
top 20% are returned as keywords.

Document Ranking

Document ranking is a central part of Compass, as mentioned previously it is the second step of the ad-hoc retrieval problem and is the area Compass tackles most directly. This process aims to promote those documents from the result set that are most similar to the source document. The method used is the *Vector Space Model* [1] with vector similarity being measured using the *Cosine measure*. Term frequencies in the vectors are dampened using $f(tf) = \sqrt{tf}$ where tf = normalized term frequency.

The vector space model is one of the most widely used models for ad-hoc retrieval. In the vector space model documents are represented as a high dimensional space in which each dimension corresponds to a word in the document, therefore any collection of words from that document (such as a set of keywords) represents a vector in that space. The magnitude of each component of the vector is given by the square root of the normalized frequency of occurrence of that word. The square root serves to dampen term frequencies because although more occurrences of a word indicate higher importance it does not represent as much relative importance as the undampened count would suggest [2].

In Compass the vector representing the source document is created using all the words of the document with the *stop words removed*. This difference from standard vector space similarity methods eliminates the impact of noise words in calculating similarity with documents. This modification has shown to be no worse than the standard version, as documents that are the same as the source document immediately move to the top of the list. A similar vector (using the only words shared with *source* document minus those in the stop word list) is created for each document in the result set. These vectors are plotted against that of the source document and similarity computed using the cosine measure. The cosine measure is a common measure of vector similarity and can be visualized as measuring the angle between a pair of vectors. It is computed as follows [2]

$$\cos(\vec{q}, \vec{d}) = \frac{(\sum_{i=1}^n q_i d_i)}{(\sqrt{\sum_{i=1}^n q_i^2} \sqrt{\sum_{i=1}^n d_i^2})}$$

where: \vec{q}, \vec{d} are n-dimensional vectors in real valued space.

The numerator represents the vector product of the two vectors and the denominator the Euclidean length of the vectors and scales for the individual magnitude of the two vectors.

In this manner documents with a higher cosine (and therefore smaller angle) to the source document, and presumably more relevant with respect to the users search goals, are moved higher up in the list.

Results & Discussion

The measure used to evaluate the results sets generated by Compass against those generated by the search engine is introduced in [2] and is known as *uninterpolated average precision*. This measure is an effective measure of how well ordered a result set is. Considering the result set as a list, it works going down from top to bottom, by measuring precision (number of relevant document / total number of documents so far) at each occurrence of a relevant document, after which these numbers are averaged. Consequently a result set which ranks all the relevant documents before the irrelevant documents will have a score of 1 no matter how many (or how few) relevant results actually exist. A simple measure of overall precision would not be useful because no matter what the ordering the result would be the same as both sets contain the same documents (thus yielding the same overall precision). Below are the results for two searches, the relevance column is marked 'x' for documents that are on topic and 'o' for off topic documents, for each occurrence of a relevant document the precision at that point in the result set is shown in the precision column. Scores range from 0 to 1 with a higher score being better. The Yahoo! search engine was used in carrying out these tests.

Experiment 1

The search query is “*mole family*”. The document and selection used as well as the full contents of the result set can be found in Appendix B-1. The overall precision for both result sets is 0.3.

Yahoo! Result Set			Compass Result Set	
Position	Relevance	Precision	Relevance	Precision
1	x	1.000	x	1.000
2	x	1.000	x	1.000
3	o		o	
4	o		o	
5	o		x	0.600
6	o		x	0.666
7	x	0.426	x	0.714
8	o		x	0.750
9	o		x	0.777
10	o		x	0.800
11	o		o	
12	x	0.333	x	
13	o		x	0.692
14	x	0.357	o	
15	x	0.400	o	
16	o		o	
17	o		o	
18	o		o	
19	o		o	
20	x	0.350	o	
21	x	0.381	o	
22	o		o	
23	o		o	
24	o		o	
25	o		o	
26	o		o	
27	o		o	
28	o		o	
29	o		o	
30	x	0.300	o	
uninterpolated average precision		0.500	0.778	

Experiment 2

The search query is “*computational linguistics discipline computer science aspects human language cognitive sciences intelligence aiming models applied*”. The document and selection used as well as the full contents of the result set can be found in Appendix B-2. The overall precision for both results set is 0.4.

Yahoo! result set			Compass result set	
Position	Relevance	Precision	Relevance	Precision
1	x	1.000	x	1.000
2	x	1.000	x	1.000
3	x	1.000	x	1.000
4	x	1.000	x	1.000
5	o		x	1.000
6	o		x	1.000
7	o	0.625	o	
8	x		x	0.875
9	x	0.666	x	0.888
10	x	0.700	x	0.909
11	o		x	0.917
12	x	0.666	o	
13	o		o	
14	o		x	0.786
15	o		o	
16	o		o	
17	o		o	
18	o		o	
19	x	0.474	x	0.632
20	o		o	
21	x	0.476	o	
22	o		o	
23	x	0.478	o	
24	o		o	
25	x	0.480	o	
26	o		o	
27	o		o	
28	o		o	
29	o		o	
30	o		o	
uninterpolated average precision		0.714	0.917	

The results clearly show the re-ranking as being effective, the first experiment suffers from low overall precision probably due to the fact that the word 'mole' is a homonym. Thus the result set has references to documents relating to the different senses of the word. However the re-ranking applied by compass is largely able to push down those documents not pertaining to the small furry mammal, resulting in much higher quality result set. Even in the second experiment where we would expect a search term as specific as the one that was used to return a fairly precise result set, the overall precision is less than half, and the default ranking provided by Yahoo! is considerably worse than that created by Compass.

It should also be evident from the definition of uninterpolated average precision that the uninterpolated average precision cannot be improved for the Yahoo! result set by increasing the number of documents considered. No matter how much larger the new result set, any new relevant document will be further down than the lowest relevant document in the smaller result set and thus cannot not give the an improved score. However the uninterpolated average precision of Compass can increase given more documents. In a larger result set a relevant document could replace one of the false positives (off topic documents high in the result set) and result in better score. It should also be noted that as the result set gets larger, recall improves

therefore Compass is able to ameliorate both precision and recall of a result set simultaneously.

Conclusion

The results presented above clearly show that Compass is effective in its original goal, to ameliorate search results with respect to user search goals as described by the document they are viewing. In the author's opinion the use of source documents in describing search goals is a useful method and the search style afforded by Compass (directly from the document) is increasingly being used in applications such as web browsers (both Mozilla Firefox and Opera have features that allow you to make a selection and via right click, send the selection to a search engine). In developing and testing this application it has shown itself to be useful in disambiguating between word senses (as in the Mole example above) as well as semantic context such as, in the second example, separating psycholinguistics from computational linguistics as well as separating Canadian from American political news (although not shown in this report, Compass was able to separate Canadian from American political news even when the search terms were generic terms relating to election). The techniques used scale well to large numbers of documents and are not restricted by language nor domain.

APPENDIX A

Stop Word List

a a's able about above according accordingly across actually after afterwards again against ain't all allow allows almost alone along already also although always am among amongst an and another any anybody anyhow anyone anything anyway anyways anywhere apart appear appreciate appropriate are aren't around as aside ask asking associated at available away awfully be became because become becomes becoming been before beforehand behind being believe below beside besides best better between beyond both brief but by c'mon c's came can can't cannot cant cause causes certain certainly changes clearly co com come comes concerning consequently consider considering contain containing contains corresponding could couldn't course currently definitely described despite did didn't different do does doesn't doing don't done down downwards during each edu eg eight either else elsewhere enough entirely especially et etc even ever every everybody everyone everything everywhere ex exactly example except far few fifth first five followed following follows for former formerly forth four from further furthermore get gets getting given gives go goes going gone got gotten greetings had hadn't happens hardly has hasn't have haven't having he he's hello help hence her here here's hereafter hereby herein hereupon hers herself hi him himself his hither hopefully how howbeit however i i'd i'll i'm i've ie if ignored immediate in inasmuch inc indeed indicate indicated indicates inner insofar instead into inward is isn't it it'd it'll it's its itself just keep keeps kept know knows known last lately later latter latterly least less lest let let's like liked likely little look looking looks ltd mainly many may maybe me mean meanwhile merely might more moreover most mostly much must my myself name namely nd

near nearly necessary need needs neither never nevertheless new
next nine no nobody non none noone nor normally not nothing novel
now nowhere obviously of off often oh ok okay old on once one ones
only onto or other others otherwise ought our ours ourselves out
outside over overall own particular particularly per perhaps placed
please plus possible presumably probably provides que quite qv
rather rd re really reasonably regarding regardless regards relatively
respectively right said same saw say saying says second secondly see
seeing seem seemed seeming seems seen self selves sensible sent
serious seriously seven several shall she should shouldn't since six so
some somebody somehow someone something sometime sometimes
somewhat somewhere soon sorry specified specify specifying still sub
such sup sure t's take taken tell tends th than thank thanks thanx that
that's thats the their theirs them themselves then thence there there's
thereafter thereby therefore therein theres thereupon these they
they'd they'll they're they've think third this thorough thoroughly
those though three through throughout thru thus to together too took
toward towards tried tries truly try trying twice two un under
unfortunately unless unlikely until unto up upon us use used useful
uses using usually value various very via viz vs want wants was wasn't
way we we'd we'll we're we've welcome well went were weren't what
what's whatever when whence whenever where where's whereafter
whereas whereby wherein whereupon wherever whether which while
whither who who's whoever whole whom whose why will willing wish
with within without won't wonder would would wouldn't yes yet you
you'd you'll you're you've your yours yourself yourselves zero

APPENDIX B – 1

The document used – Moles (Animal) from Wikipedia
http://en.wikipedia.org/wiki/Mole_%28animal%29

Moles are members of the family (Talpidae) of mammals in the order Insectivora that live underground, burrowing holes. Some species are aquatic or semi-aquatic. They have cylindrical bodies covered in fur with small or covered eyes; the ears are generally not visible. They feed on small invertebrate animals living under ground. Moles can be found in North America, Europe and Asia.

The family is divided into three subfamilies:

- * Desmaninae (desmans)
- * Talpinae (moles)
- * Uropsilinae (shrew moles)

Male moles are called boars, females are called sows. A group of moles is called a labour.

Contents

Diet

Moles eat worms. They maintain a fresh supply of worms in underground chambers. The mole immobilizes the worms by biting their heads off.

http://www.eurekalert.org/pub_releases/2005-02/vu-mm013105.php
describes how the Star-nosed Mole can detect, catch and eat food faster than the human eye can follow (under 300 milliseconds).

Pest

Moles are considered to be an agricultural pest in some countries , while in others such as Germany they are a protected species. Problems caused are cited as contamination of silage with soil particles making it unpalatable to animals, the covering of pasture with fresh soil reducing its size and yield , damage to agricultural machinery by the exposure of stones, damage to young plants through disturbance of the soil, weed invasion of pasture through exposure of fresh tilled soil, and damage to drainage systems and watercourses. Other species such as weasels and voles may use mole tunnels to gain access to enclosed areas or plant roots.

Moles that burrow in the lawns of humans can disturb the earth, raising molehills and causing enough aesthetic problems to be considered as pests. They do however benefit the soil by aerating and tilling it, adding to its fertility.

They are controlled with traps and poisons such as calcium carbide and strychnine. .

Similar Animals

Other similar animals are found in family Chrysochloridae, the golden moles, also in order Insectivora, and family Notoryctidae, the marsupial moles, which are not related to true moles.

There are also similar-looking but herbivorous rodents called mole-rats that enjoy a similar life-style and are commonly called "moles", although, unlike mole-rats, no species of true mole is known to be eusocial.

Trivia

- * Many types of earth-burrowing machines take their names from moles.
- * Dogs have been known to dig moles out of the ground and bite them to death.

* One of British comedian Jasper Carrott's sketches involves him trying to get rid of a mole that had moved into to his garden, at one point, as he described, trying to use a shotgun, a swirling chair, and a torch.

Common Myth

Moles are often said to be blind but this is incorrect. The mole has eyes and ears but they are incredibly small so they do not become filled with earth when digging. The eyes are the size of pinheads and are normally kept closed. Moles' eyesight is poor. They can see movement and tell light from dark, but cannot distinguish colours. The ears are about 5 centimetres long and hidden in the fur. There is an area of bare pink skin on the snout that is very sensitive to touch. Sensitive pimples cover the surface, detecting tiny movements, temperature change, prey and other moles through scent. The mole's long canine teeth are sharp and pierce the hard outer skeleton of insect prey. The tail is held upwards to feel its way when running backwards in tunnels. Mole fur is very fine and velvety and usually black, it can point in any direction with ease, so is specially adapted for life in tunnels.

The selection from which the search query was generated (from first line): Moles are members of the family (Talpidae)

Yahoo Results

Search terms: moles family

Date Mon Apr 24 17:42:38 EDT 2006

Moles (Family Talpidae)

<http://www.ku.edu/~mammals/mole.html>

Description Eastern Mole. *Scalopus aquaticus* (Linnaeus) Description: Eastern moles are related to shrews, and of the four kinds of insectivores in Kansas, they are best adapted to underground life. ... Surface tunnels of eastern moles are characterized by continuous ...

Mole, Mammals, Mole Picture, Mammal Pictures, Catalog, Encyclopedia

<http://www.4to40.com/earth/geography/htm/mammalsindex.asp?counter=103>

Description ... The golden moles, family Chrysochloridae in the order Insectivora, are an African group of 15 to 20 ... The MARSUPIAL moles, family Notoryctidae in the order Marsupalia, comprise one or ...

The Emmett Moles Family Information

<http://showcase.netins.net/web/marjned/emmett.html>

Description The Emmett Moles Family Information. Emmett Moles Family Information. Emmett Moles was born on 13 Apr 1892 in Carroll County, Virginia. His wife, Pearl Ellen Wankel, was born on 6 Jul 1891. They were married 17 Dec 1914.

Family of Spies: The John Walker Jr. Spy Case

http://www.crimelibrary.com/terrorists_spies/spies/walker/1.html

Description John Walker, with two recruits, sold secret information to the Russians for 18 years, from the Crime Library ... Death in the Family. Celebrity Crimes ... Mob Bossess. Crime Family Epics. Unique Gang Organizations ... You are in: TERRORISTS, SPIES & ASSASSINS/SPIES. FAMILY OF SPIES: THE JOHN WALKER JR ...

Medical Library Search

http://www.medem.com/search/article_display.cfm?path=n:&mstr=/ZZZD4SPC2BC.html&soc=JAMA/Archives&srch_typ=NAV_SERCH

Description Medical Library. Real Patient Cases from JAMA 'Clinical Crossroads' Case History: Multiple Moles. Ms G has had moles for as long as she can remember. ... history of skin cancer. Her family history revealed a first cousin with multiple moles, but was otherwise unremarkable ...

Mammals of Kansas

<http://www.ku.edu/~mammals/list.html>

Description ... Moles (Family Talpidae) Eastern Mole *Scalopus aquaticus*. Common Bats (Family Vespertilionidae) Pallid Bat *Antrozous pallidus* ...

Information about Moles

<http://tim.rawle.org/moles>

Description Moles are members of a family of small burrowing insectivores found in Europe, Asia and North and Central America. They have velvety, typically dark (except for the golden moles) fur and forearms specialized for digging. ... resistance against the water. Golden moles (family Chrysochloridae) have strong forelimbs each with four ... Tail. Golden moles (family Chrysochloridae), have no external sign of a tail ...

Obituary of Anson Kelly Moles (1874-1919), and Others
[http://home1.gte.net/gentzler/Moles_Anson_Kelly_\(obit\).htm](http://home1.gte.net/gentzler/Moles_Anson_Kelly_(obit).htm)

Description "KILLED BY R.I. TRAIN. A.K. Moles Victim of Accident. Drove Onto Nelson Line Tracks. in Front of Engine" "A.K. ... "A.K. Moles was killed Tuesday morning at the Rock Island railroad crossing, west of Fairbury, near ... that covered wagon. The Moles family arrived at their homestead site (located ...

Skin Moles
<http://www.thecountrydoctor.com/prescriptions/skinmoles.htm>

Description Skin moles and growths of all kinds get into the news now and then. This time it is because of the depletion of the earth's ozone layer. The problem happens when we earth-people pollute the air we breathe with certain toxic chemicals. ... hear of more than one member in a family with multiple moles: siblings; parents; grandparents; cousins; and so on ...

Walker Moles Family Information
<http://showcase.netins.net/web/marjned/walker.html>

Description ... Walker Moles Family Information. Walker Moles was born on 7 Jul 1864 in Floyd County, Virginia ... Cassell B. Moles was born on 20 Aug 1885 in Carroll County, Virginia ...

Georgia's Mammals
<http://www.gwf.org/commonmammals.htm>

Description ... SHREWS & MOLES (Order Insectivora) Moles (Family Talpidae) Eastern Mole *Scalopus aquaticus* ... Beaked Whales (Family Ziphiidae) Dense-beaked Whale *Mesoplodon densirostris* ...

Moles Management Guidelines--UC IPM
<http://www.ipm.ucdavis.edu/PMG/PESTNOTES/pn74115.html>

Description UC home and landscape guidelines for control of moles. (Published: 5/04) ... it is not part of the rodent family. In California, moles inhabit the Sierra Nevada, the coastal ... Moles live almost entirely underground in a vast network of interconnecting tunnels ...

Poole Dermatology and The Laser Center of New Orleans- Dermatology
<http://www.pooledermatology.com/dermatology.html>

Description ... Moles formally known as nevi, are quite common in all ages ... age, with many adults having 30-40 moles. Family history of moles and frequent sun exposure both contribute to ...

NSiS: Florida Wildlife - Moles
<http://www.nsis.org/wildlife/mamm/mole.html>

Description ... Moles. Moles (family Talpidae), like all insectivores, feed mainly on insects but also eat other small ... small ears, and long, pointed, flexible snouts. Moles have wide front feet ...

Mole (animal) - Wikipedia, the free encyclopedia
[http://en.wikipedia.org/wiki/Mole_\(animal\)](http://en.wikipedia.org/wiki/Mole_(animal))

Description Mole (animal) From Wikipedia, the free encyclopedia Mole Genera Moles are members of the family (Talpidae) of mammals in the order Insectivora that live underground, burrowing holes. Some species are aquatic or semi-aquatic. ... animals living under ground. Moles can be found in North America, Europe ...

Moles
<http://parentcenter.babycenter.com/refcap/bigkid/ghealth/grash/68480.html>

Description What those marks on your child's skin are and whether you need to worry. ... Do moles change over time? Yes. Moles generally go through a life cycle of about 50 years of ... as the Best Family/Parenting site. Please cast your vote for BabyCenter ...

Mole (skin marking) - Wikipedia, the free encyclopedia
[http://en.wikipedia.org/wiki/Mole_\(skin_marking\)](http://en.wikipedia.org/wiki/Mole_(skin_marking))

Description Mole (skin marking) From Wikipedia, the free encyclopedia Melanocytic nevus Melanocytic naevus A mole or melanocytic naevus is a small, dark spot on the skin. ... are a member of the family of skin lesions known as naevi. Moles may either be present at birth ... According to some people, moles may enhance or detract from beauty ...

Brigham And Women's Hospital Health Information - Moles
<http://healthgate.partners.org/browsing/browseContent.asp?fileName=11669.xml&title=Moles>

Description ... Moles present at birth (this increases the likelihood of more moles later in life) Family members who have moles. Excessive exposure to ...

Atypical (dysplastic) Moles
http://www.aocd.org/skin/dermatologic_diseases/atypical_moles.html

Description ... Atypical (dysplastic) Moles. For years, doctors have debated the risk of developing melanoma in people with ... a great many atypical moles and several family members have had melanoma ...

Moles
<http://www.epestsupply.com/moles.htm>

Description ... there are some differences between the various species of moles, most moles have similar biology and behavior ... having to contend with a family of moles later. For successful trapping ...

Moles
<http://www.dpest.com/moles.htm>

Description ... Moles are from the family Talpidae which includes moles, desmans, and shrew-moles. The North American moles or New World moles are divided into ...

Moles Family Services (Ferndale, WA) | Company Reports

http://rdre1.yahoo.com/click?u=http://www.manta.com/comsite5/bin/pddnb_compan y.pl%3Fpdlanding%3D1%26referid%3D3550%26id%3Dcd5smjv&y=0437A9F7F68A7B8C58&i=482&c=24105&q=02%5ESSHPM%5BL7rpszl%3Fy~rvsf%3F6%3FYVSKZM7ypmr~k%22%22wkrs6&e=utf-8&r=21&d=wow~WBSV-en-us&n=UJF4G7GH8D6K9IG8&s=10&t=&m=444D463E&x=05CC3A6C922E507C009482B86FD18F7282

Description Moles Family Services (Ferndale, WA) company info from D&B. Purchase and instantly access business information, credit ratings, company history, location, financial data, and more...

Moles, what they are and how to get rid of them

<http://www.no-moles.com/moles.htm>

Description Natural herbal remedy removes moles, warts, skin tags and benign skin cancers, naturally. ... Additional information about moles. Moles are spots on the skin ... when scientists observed that members of a melanoma-prone family had numerous large, abnormal moles on their bodies ...

Prevention and Early Detection of Malignant Melanoma - November 15, 2000 - American Family Physician

<http://www.aafp.org/afp/20001115/2277.html>

Description ... the problem of malignant melanoma, family physicians should emphasize primary prevention ... in childhood), atypical moles, large numbers of acquired nevi, a family history of melanoma ...

Moles

<http://www.babycenter.com/refcap/baby/babyills/babyrash/1435115.html>

Description You seem to have disabled JavaScript. To take full advantage of our site and functionality, we recommend you enable Javascript. How? By the BabyCenter editorial staff. Do moles change over time? Yes. ... Do moles change over time? Yes. Moles generally go through a life cycle of about 50 years of ... as the Best Family/Parenting site. Please cast your vote for BabyCenter ...

Moles

<http://healthlibrary.epnet.com/GetContent.aspx?token=af362d97-4f80-4453-a175-02cc6220a387&chunkiid=11669>

Description (Skin Nevi) Definition. Moles are small growths on the skin that develop from pigment-producing cells in the skin called melanocytes. They typically appear as dark brown spots on the skin that are either flat or raised. ... Moles present at birth (this increases the likelihood of more moles later in life) Family members who have moles. Excessive exposure to ...

Moles Family Funeral Homes & Crematory

<http://www.molesfamilyservices.com/>

Description For four generations the Moles family has been the trusted provider of memorial care to the families of Whatcom County.

Screening Moles for Cancer

<http://www.webmd.com/content/article/117/112595.htm>

Description When it comes to the health of your skin, it's a good idea to be proactive and keep an eye out for dangerous moles. Moles can be linked to skin cancer. This is especially true if you already have moles on your body, or if you have a family ...

Family of Mariah Tope and William Moles

http://www.geocities.com/donaldkear/george02/maria_tope.htm

Description ... George TOPE and Elizabeth BOWER. Family of Mariah TOPE and William MOLES/MOULS. "Maria Tope married William Moles [note]; they moved to Marion county, Iowa, where they lived a ...

Moles and Shrews

<http://www.extension.umn.edu/distribution/naturalresources/DD1139.html>

Description Provides ways to control moles and shrews near the home. Shows how to identify moles and shrews and describes possible mole damage to yards and gardens. Describes and illustrates several types of traps used for mole and shrew control. ... Several species of moles inhabit the United States, but only the common (or Eastern) mole and the star-nosed mole ...

Compass Results

Search terms: moles family

Date Mon Apr 24 17:43:07 EDT 2006

Mole (animal) - Wikipedia, the free encyclopedia
[http://en.wikipedia.org/wiki/Mole_\(animal\)](http://en.wikipedia.org/wiki/Mole_(animal))

Description Mole (animal) From Wikipedia, the free encyclopedia Mole Genera
Moles are members of the family (Talpidae) of mammals in the order Insectivora
that live underground, burrowing holes. Some species are aquatic or semi-aquatic.
... animals living under ground. Moles can be found in North America, Europe ...
Score: 0.7330171906868602

Moles
<http://www.dpest.com/moles.htm>

Description ... Moles are from the family Talpidae which includes moles, desmans,
and shrew-moles. The North American moles or New World moles are divided into
...
Score: 0.6727138381177364

Moles, what they are and how to get rid of them
<http://www.no-moles.com/moles.htm>

Description Natural herbal remedy removes moles, warts, skin tags and benign skin
cancers, naturally. ... Additional information about moles. Moles are spots on the
skin ... when scientists observed that members of a melanoma-prone family had
numerous large, abnormal moles on their bodies ...
Score: 0.6599013078041094

Walker Moles Family Information
<http://showcase.netins.net/web/marjned/walker.html>

Description ... Walker Moles Family Information. Walker Moles was born on 7 Jul
1864 in Floyd County, Virginia ... Cassell B. Moles was born on 20 Aug 1885 in
Carroll County, Virginia ...
Score: 0.6283977280731086

Moles Management Guidelines--UC IPM
<http://www.ipm.ucdavis.edu/PMG/PESTNOTES/pn74115.html>

Description UC home and landscape guidelines for control of moles. (Published:
5/04) ... it is not part of the rodent family. In California, moles inhabit the Sierra
Nevada, the coastal ... Moles live almost entirely underground in a vast network of
interconnecting tunnels ...
Score: 0.6147376085980384

Moles

<http://www.epestsupply.com/moles.htm>

Description ... there are some differences between the various species of moles, most moles have similar biology and behavior ... having to contend with a family of moles later. For successful trapping ...

Score: 0.6140068723369189

Mole, Mammals, Mole Picture, Mammal Pictures, Catalog, Encyclopedia

<http://www.4to40.com/earth/geography/htm/mammalsindex.asp?counter=103>

Description ... The golden moles, family Chrysochloridae in the order Insectivora, are an African group of 15 to 20 ... The MARSUPIAL moles, family Notoryctidae in the order Marsupalia, comprise one or ...

Score: 0.613946208789669

Information about Moles

<http://tim.rawle.org/moles>

Description Moles are members of a family of small burrowing insectivores found in Europe, Asia and North and Central America. They have velvety, typically dark (except for the golden moles) fur and forearms specialized for digging. ... resistance against the water. Golden moles (family Chrysochloridae) have strong forelimbs each with four ... Tail. Golden moles (family Chrysochloridae), have no external sign of a tail ...

Score: 0.6002073421016578

Moles (Family Talpidae)

<http://www.ku.edu/~mammals/mole.html>

Description Eastern Mole. *Scalopus aquaticus* (Linnaeus) Description: Eastern moles are related to shrews, and of the four kinds of insectivores in Kansas, they are best adapted to underground life. ... Surface tunnels of eastern moles are characterized by continuous ...

Score: 0.579033498788714

Moles and Shrews

<http://www.extension.umn.edu/distribution/naturalresources/DD1139.html>

Description Provides ways to control moles and shrews near the home. Shows how to identify moles and shrews and describes possible mole damage to yards and gardens. Describes and illustrates several types of traps used for mole and shrew control. ... Several species of moles inhabit the United States, but only the common (or Eastern) mole and the star-nosed mole ...

Score: 0.5562288375923812

Atypical (dysplastic) Moles

http://www.aocd.org/skin/dermatologic_diseases/atypical_moles.html

Description ... Atypical (dysplastic) Moles. For years, doctors have debated the risk of developing melanoma in people with ... a great many atypical moles and several family members have had melanoma ...

Score: 0.5448837619707533

Obituary of Anson Kelly Moles (1874-1919), and Others

[http://home1.gte.net/gentzler/Moles_Anson_Kelly_\(obit\).htm](http://home1.gte.net/gentzler/Moles_Anson_Kelly_(obit).htm)

Description "KILLED BY R.I. TRAIN. A.K. Moles Victim of Accident. Drove Onto Nelson Line Tracks. in Front of Engine" "A.K. ... "A.K. Moles was killed Tuesday morning at the Rock Island railroad crossing, west of Fairbury, near ... that covered wagon. The Moles family arrived at their homestead site (located ...

Score: 0.5062868202244164

NSiS: Florida Wildlife - Moles

<http://www.nsis.org/wildlife/mamm/mole.html>

Description ... Moles. Moles (family Talpidae), like all insectivores, feed mainly on insects but also eat other small ... small ears, and long, pointed, flexible snouts. Moles have wide front feet ...

Score: 0.5053367523300817

The Emmett Moles Family Information

<http://showcase.netins.net/web/marjned/emmett.html>

Description The Emmett Moles Family Information. Emmett Moles Family Information. Emmett Moles was born on 13 Apr 1892 in Carroll County, Virginia. His wife, Pearl Ellen Wankel, was born on 6 Jul 1891. They were married 17 Dec 1914.

Score: 0.49454813265783654

Medical Library Search

http://www.medem.com/search/article_display.cfm?path=n:&mstr=/ZZZD4SPC2BC.html&soc=JAMA/Archives&srch_typ=NAV_SERCH

Description Medical Library. Real Patient Cases from JAMA 'Clinical Crossroads' Case History: Multiple Moles. Ms G has had moles for as long as she can remember. ... history of skin cancer. Her family history revealed a first cousin with multiple moles, but was otherwise unremarkable ...

Score: 0.49106341457529556

Mole (skin marking) - Wikipedia, the free encyclopedia

[http://en.wikipedia.org/wiki/Mole_\(skin_marking\)](http://en.wikipedia.org/wiki/Mole_(skin_marking))

Description Mole (skin marking) From Wikipedia, the free encyclopedia Melanocytic nevus Melanocytic naevus A mole or melanocytic naevus is a small, dark spot on the skin. ... are a member of the family of skin lesions known as naevi. Moles may either be present at birth ... According to some people, moles may enhance or detract from beauty ...

Score: 0.4769684107686237

Moles

<http://healthlibrary.epnet.com/GetContent.aspx?token=af362d97-4f80-4453-a175-02cc6220a387&chunkid=11669>

Description (Skin Nevi) Definition. Moles are small growths on the skin that develop from pigment-producing cells in the skin called melanocytes. They typically appear as dark brown spots on the skin that are either flat or raised. ... Moles present at birth (this increases the likelihood of more moles later in life) Family members who have moles. Excessive exposure to ...

Score: 0.46051077129096535

Skin Moles

<http://www.thecountrydoctor.com/prescriptions/skinmoles.htm>

Description Skin moles and growths of all kinds get into the news now and then. This time it is because of the depletion of the earth's ozone layer. The problem happens when we earth-people pollute the air we breathe with certain toxic chemicals. ... hear of more than one member in a family with multiple moles: siblings; parents; grandparents; cousins; and so on ...

Score: 0.4458170567066016

Family of Mariah Tope and William Moles

http://www.geocities.com/donaldkear/george02/maria_tope.htm

Description ... George TOPE and Elizabeth BOWER. Family of Mariah TOPE and William MOLES/MOULS. "Maria Tope married William Moles [note]; they moved to Marion county, Iowa, where they lived a ...

Score: 0.4115891677135376

Screening Moles for Cancer

<http://www.webmd.com/content/article/117/112595.htm>

Description When it comes to the health of your skin, it's a good idea to be proactive and keep an eye out for dangerous moles. Moles can be linked to skin cancer. This is especially true if you already have moles on your body, or if you have a family ...

Score: 0.40024271261171734

Brigham And Women's Hospital Health Information - Moles

<http://healthgate.partners.org/browsing/browseContent.asp?fileName=11669.xml&title=Moles>

Description ... Moles present at birth (this increases the likelihood of more moles later in life) Family members who have moles. Excessive exposure to ...

Score: 0.39606150256272343

Moles

<http://parentcenter.babycenter.com/refcap/bigkid/ghealth/grash/68480.html>

Description What those marks on your child's skin are and whether you need to worry. ... Do moles change over time? Yes. Moles generally go through a life cycle of about 50 years of ... as the Best Family/Parenting site. Please cast your vote for BabyCenter ...
Score: 0.39260777975872196

Poole Dermatology and The Laser Center of New Orleans- Dermatology
<http://www.pooledermatology.com/dermatology.html>

Description ... Moles formally known as nevi, are quite common in all ages ... age, with many adults having 30-40 moles. Family history of moles and frequent sun exposure both contribute to ...
Score: 0.3905456361017487

Moles
<http://www.babycenter.com/refcap/baby/babyills/babyrash/1435115.html>

Description You seem to have disabled JavaScript. To take full advantage of our site and functionality, we recommend you enable Javascript. How? By the BabyCenter editorial staff. Do moles change over time? Yes. ... Do moles change over time? Yes. Moles generally go through a life cycle of about 50 years of ... as the Best Family/Parenting site. Please cast your vote for BabyCenter ...
Score: 0.3868362924852603

Prevention and Early Detection of Malignant Melanoma - November 15, 2000 - American Family Physician
<http://www.aafp.org/afp/20001115/2277.html>

Description ... the problem of malignant melanoma, family physicians should emphasize primary prevention ... in childhood), atypical moles, large numbers of acquired nevi, a family history of melanoma ...
Score: 0.3777929590942446

Mammals of Kansas
<http://www.ku.edu/~mammals/list.html>

Description ... Moles (Family Talpidae) Eastern Mole Scalopus aquaticus. Common Bats (Family Vespertilionidae) Pallid Bat Antrozous pallidus ...
Score: 0.3625607283327402

Moles Family Services (Ferndale, WA) | Company Reports
http://rdre1.yahoo.com/click?u=http://www.manta.com/comsite5/bin/pddnb_company.pl%3Fpdlanding%3D1%26referid%3D3550%26id%3Dcd5smjv&y=0437A9F7F68A7B8C58&i=482&c=24105&q=02%5ESSHPM%5BL7rpszl%3Fy~rvsf%3F6%3FYVSKZM7ypmr~k%22%22wkrs6&e=utf-8&r=21&d=wow~WBSV-en-us&n=UJF4G7GH8D6K9IG8&s=10&t=&m=444D463E&x=05CC3A6C922E507C009482B86FD18F7282

Description Moles Family Services (Ferndale, WA) company info from D&B. Purchase and instantly access business information, credit ratings, company history, location, financial data, and more...
Score: 0.3222899670762006

Moles Family Funeral Homes & Crematory
<http://www.molesfamilyservices.com/>

Description For four generations the Moles family has been the trusted provider of memorial care to the families of Whatcom County.
Score: 0.2934561121656845

Georgia's Mammals
<http://www.gwf.org/commonmammals.htm>

Description ... SHREWS & MOLES (Order Insectivora) Moles (Family Talpidae) Eastern Mole *Scalopus aquaticus* ... Beaked Whales (Family Ziphiidae) Dense-beaked Whale *Mesoplodon densirostris* ...
Score: 0.24049469095589474

Family of Spies: The John Walker Jr. Spy Case
http://www.crimelibrary.com/terrorists_spies/spies/walker/1.html

Description John Walker, with two recruits, sold secret information to the Russians for 18 years, from the Crime Library ... Death in the Family. Celebrity Crimes ... Mob Bossess. Crime Family Epics. Unique Gang Organizations ... You are in: TERRORISTS, SPIES & ASSASSINS/SPIES. FAMILY OF SPIES: THE JOHN WALKER JR ...
Score: 0.202303702151397

APPENDIX B – 2

The document used – What is Computational Linguistics? -Hans Uszkoreit

http://www.coli.uni-saarland.de/~hansu/what_is_cl.html

WHAT IS COMPUTATIONAL LINGUISTICS?

Computational linguistics (CL) is a discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences and overlaps with the field of artificial intelligence (AI), a branch of computer science aiming at computational models of human cognition. Computational linguistics has applied and theoretical components.

Human language is a most exciting and demanding puzzle.

Theoretical CL takes up issues in theoretical linguistics and cognitive science. It deals with formal theories about the linguistic knowledge that a human needs for generating and understanding language. Today these theories have reached a degree of complexity that can only be managed by employing computers.

Computational linguists develop formal models simulating aspects of the human language faculty and implement them as computer programmes. These programmes constitute the basis for the evaluation and further development of the theories. In addition to linguistic theories, findings from cognitive psychology play a major role in simulating linguistic competence. Within psychology, it is mainly the area of psycholinguistics that examines the cognitive processes constituting human language use. The relevance of computational modelling for psycholinguistic research is reflected in the emergence of a new subdiscipline: computational psycholinguistics.

We teach computers to communicate with people.

Applied CL focusses on the practical outcome of modelling human language use. The methods, techniques, tools and applications in this area are often subsumed under the term language engineering or (human) language technology. Although existing CL systems are far from achieving human ability, they have numerous

possible applications. The goal is to create software products that have some knowledge of human language. Such products are going to change our lives. They are urgently needed for improving human-machine interaction since the main obstacle in the interaction between human and computer is a communication problem. Today's computers do not understand our language but computer languages are difficult to learn and do not correspond to the structure of human thought. Even if the language the machine understands and its domain of discourse are very restricted, the use of human language can increase the acceptance of software and the productivity of its users.

Friendly software should listen and speak.

Natural language interfaces enable the user to communicate with the computer in French, English, German, or another human language. Some applications of such interfaces are database queries, information retrieval from texts, so-called expert systems, and robot control. Current advances in the recognition of spoken language improve the usability of many types of natural language systems.

Communication with computers using spoken language will have a lasting impact upon the work environment, completely new areas of application for information technology will open up. However, spoken language needs to be combined with other modes of communication such as pointing with mouse or finger. If such multimodal communication is finally embedded in an effective general model of cooperation, we have succeeded in turning the machine into a partner.

Machines can also help people communicate with each other.

Much older than communication problems between human beings and machines are those between people with different mother tongues. One of the original aims of applied computational linguistics has always been fully automatic translation between human languages. From bitter experience scientists have realized that they are still far away from achieving the ambitious goal of translating unrestricted texts. Nevertheless computational linguists have created software systems that simplify the work of human translators and clearly improve their productivity. Less than perfect automatic translations can also be of great help to information seekers who have to search through large amounts of texts in foreign languages.

Language is the fabric of the web.

The rapid growth of the Internet/WWW and the emergence of the information society poses exciting new challenges to language technology. Although the new media combine text, graphics, sound and movies, the whole world of multimedia information can only be structured, indexed and navigated through language. For browsing, navigating, filtering and processing the information on the web, we need software that can get at the contents of documents. Language technology for content management is a necessary precondition for turning the wealth of digital information into collective knowledge. The increasing multilinguality of the web constitutes an additional challenge for our discipline. The global web can only be mastered with the help of multilingual tools for indexing and navigating. Systems for crosslingual information and knowledge management will surmount language barriers for e-commerce, education and international cooperation.

Our discipline combines ambitious visions and realistic applications.

We still do not know very well how people produce and comprehend language. Yet our understanding of the intricate mechanisms that underly human language processing keeps growing. Modelling such mechanisms on a computer also helps us to discover and formally describe hidden properties of human language that are relevant for any kind of language processing including many useful software applications. Our long term goal is the deep understanding of human language and powerful intelligent linguistic applications. However, even today's language technologies full of clever short cuts and shallow processing techniques can be turned into badly needed software products.

The young field's appeal is its diversity.

For many students and practitioners of computational linguistics the special attraction of the discipline is the combination of expertise from the humanities, natural and behavioural sciences, and engineering. Scientific approaches and practical techniques come from linguistics, computer science, psychology, and mathematics. At some universities the subject is taught in computer science at others it belongs to linguistics or cognitive science. In addition there is a small but growing number of programs and departments dedicated solely to computational linguistics.

The selection from which the search query was generated (first paragraph): Computational linguistics (CL) is a discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences and overlaps with the field of artificial intelligence (AI), a branch of computer science aiming at computational models of human cognition. Computational linguistics has applied and theoretical components

Yahoo Results

*Search terms: computational linguistics discipline computer science aspects
human language cognitive sciences intelligence aiming models applied*

Date Tue Apr 25 20:04:00 EDT 2006

Introductory

<http://www.msu.edu/course/lin/475/notes/hale-notes002.html>

Description 1.1 What is Computational Linguistics? ... psychological models of ... modeling human linguistic ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...

CL Intro Text

http://www.coli.uni-saarland.de/~hansu/what_is_cl.html

Description WHAT IS COMPUTATIONALLINGUISTICS? ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... linguistics has applied ...

NLP Reading

<http://www.infm.ulst.ac.uk/~MAEVE/nlp/reading.html>

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Littera:

<http://littera.deusto.es/prof/abaitua/hlt/hlt0304/Computational%20Linguistics>

Description Computational Linguistics is a subfield of Linguistics in which logical modelling of natural language from an computational perspective is central. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... linguistics has applied ...

Psychology - Wikipedia, the free encyclopedia

<http://en.wikipedia.org/wiki/Psychology>

Description ... models of human ... computer science and neuroscience) in the quest to understand the mind, the umbrella discipline of cognitive ... models from cognitive psychology are widely applied ...

News Indexed by Topic - COGNITIVE SCIENCE ARCHIVE

<https://www.aaii.org/AITopics/newstopics/cognitive1.html>

Description AI Topics provides basic, understandable information and helpful resources concerning artificial intelligence, with an emphasis on material available online. ... computational sciences is the understanding of the mechanisms of the human ... new discipline of cognitive science, the ... computer." >>> Cognitive Science, Nature of Intelligence ...

ACT | Departments | Philosophy and Social Sciences | Course Descriptions

http://www.anatolia.edu.gr/act/academics/departments/philsoc/course_descriptions.shtml

Description Course Descriptions. Philosophy. Philosophy 101: Introduction to Philosophy and Critical Reasoning. The primary aim of this course is to train students in the skills required for critical analysis of discourse. ... Cognitive Science. Researchers in psychology, linguistics, computer ... aspects of human intelligence ... sciences and one of anthropology's distinctive characteristics as a discipline ...

Linguist List - Web Resource Listings
<http://www.ling.ed.ac.uk/linguist/sp/Projects.html>

Description ... the Cognitive Science Building on ... Computational Linguistics: Department of Computer Science, Technion, Israel Institute of Technology. Lakota Language ... computational models of human ...

Natural Language Processing FAQ
http://icl.pku.edu.cn/icl_members/member/zwd/NLP/OVERVIEW/NLP-FAQ.htm

Description Natural Language Processing FAQ. From: radev@tune.cs.columbia.edu (Dragomir R. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Newsgroup FAQ's: comp.ai/Natural Language Processing FAQ
http://www.newsville.com/cgi-bin/getfaq?file=comp.ai/Natural_Language_Processing_FAQ

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Newsgroup FAQ's: comp.ai.nat-lang/Natural Language Processing FAQ
http://www.newsville.com/cgi-bin/getfaq?file=comp.ai.nat-lang/Natural_Language_Processing_FAQ

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Computer Laboratory - The Blue Book - University Teaching Officers
<http://www.cl.cam.ac.uk/DeptInfo/Overview/node4.html>

Description ... cognitive science and ... computational and theoretical linguistics and automated speech and language processing. He directs and is heavily involved in the teaching of the MPhil in Computer ...

Natural Language Processing FAQ
<http://omicron.felk.cvut.cz/FAQ/articles/a1248.html>

Description ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Psychology/rewrite - Wikipedia, the free encyclopedia
<http://en.wikipedia.org/wiki/Psychology/rewrite>

Description Psychology/rewrite From Wikipedia, the free encyclopedia This article is in the process of being merged into Psychology, and may be outdated. ... human psychology, known collectively as cognitive ... linguistics, computer science ... applied universally to psychology, as some areas within the discipline, such as cognitive science ...

TopNavBar

http://www.olats.org/space/seti/2003/mono_index.html

Description ... Extraterrestrial Intelligence (SETI ... Human-Computer Interaction (HCI). He joined the faculty at the University of Birmingham in 1986 to research and teach Cognitive Science (linguistics ... computational models of ...

Year 2006 News Archive: February

<http://www.aaai.org/aitopics/html/archvG2.html>

Description AI Topics provides basic, understandable information and helpful resources concerning artificial intelligence, with an emphasis on material available online. ... Models Computer on ... human terms. Moss said he'd also like to initiate collaborations with the Computer Science and Artificial Intelligence ...

Présentation de la journée.

<http://www.olats.org/space/seti/2003/space13Avr2003.html>

Description Présentation. On March 23-24, 2003, the second in a series of international workshops on interstellar message design will be held in Paris. ... Human-Computer Interaction (HCI). He joined the faculty at the University of Birmingham in 1986 to research and teach Cognitive Science (linguistics ... computational models of ...

Psychology - Wiki

<http://www.yikipedia.org/mediawiki/index.php/Psychology>

Description Psychology. From Wiki. Psychology. Areas. Lists ... studying human psychology ... Cognitive science again considers the "mind" as a subject for investigation, using the tools of evolutionary psychology, linguistics, computer ... computational models to ...

IBRI Research Report #31 - BEYOND THE SHADOW OF A DOUBT: Logical Deduction and the Reasoning Process

<http://www.ibri.org/RRs/RR031/31shadow.html>

Description ... The science of artificial intelligence as applied to image ... a human intelligence in all measurable aspects, then for ... and linguistics, and other branches of cognitive science'). What ...

Computers and the English language

<http://elex.amu.edu.pl/~przemka/Lectures/Lecture02pk.htm>

Description ... computer-human interaction ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... computational models of ...

MITECS: Culture, Cognition, and Evolution

<http://www.cs.buffalo.edu/pub/WWW/faculty/rapaport/575/FO1/mycultureintro.html>

Description ... cognitive sciences focuses ... more human-like ... human mind as essentially a general intelligence. Arguments and evidence from evolutionary theory, developmental psychology, linguistics, and one approach in cognitive ...

My Research Area

<http://www.cfilt.iitb.ac.in/~debasri/research.html>

Description Research Interests. My Research Area is Natural Language Processing. This also known as Computational Linguistics. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...

Formal Methods

<http://nora.hd.uib.no/AcoHum/fm/fm-chapter-final.html>

Description ... with Applied Computing ... Cognitive and Computing Sciences, offers humanities related subjects including Linguistics (including computer linguistics), philosophy, and Computer Science ... of human language ...

Linguist List - Web Resource Listings

<http://linglist.org/sp/Projects.html>

Description Fund Drive 2006 has set sail! Projects & Research Sites ... the Cognitive Science Building on ... Computational Linguistics: Department of Computer Science, Technion, Israel Institute of Technology. Lakota Language ... computational models of human ...

Machine Learning List: Vol. 9, No. 16

<http://www.ics.uci.edu/~mlearn/MList/v9/16.html>

Description ... applied methods from statistics, computer science, artificial intelligence ... complementary aspects of ... Intelligence and Cognitive Science, including Cognitive Psychology, Computational Linguistics ...

Science - Social Sciences - Linguistics - Computational Linguistics - Newsletter - News - Reviews - Education - ...

http://www.banner-net.com/Science/Social_Sciences/Linguistics/Computational_Linguistics

Description ... done aiming for ... of human communication ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...

The KLI Theory Lab - Ethology, Sociobiology

<http://www.kli.ac.at/theorylab/Areas/ETSB.html>

Description KLI Theory Lab — Area Introductions. ETSB. ETHOLOGY AND. SOCIOBIOLOGY. Even though our bodies and behaviors share many common attributes, it's far more fruitful to consider not one human nature but many. ... human behavior, aiming to ... cognitive sciences, comparative ... cognitive science, including artificial intelligence, human information processing, linguistics, mathematical models ...

EpistemeLinks: For Philosophy Resources on the Internet

<http://www.epistemelinks.com/Main/Journals.aspx?Publisher=Kluwer>

Description EpistemeLinks is a comprehensive resource for philosophy on the Internet, providing thousands of links categorized by philosopher, topic, and resource type. ... Intelligence, Cognitive Science and ... science in general and natural and human sciences ... aspects of artificial intelligence concerned with language such as computational linguistics ...

THE CHALLENGE OF SOCIOCYBERNETICS

http://uwacadweb.uwyo.edu/RED_FEATHER/chaos/006challenges.html

Description ... have applied it to the social sciences: Deutsch2 ... mental" models of their environment that act like computer subroutines ... Embodied Mind - Cognitive Science and Human Experience (3rd ...

http://www.garfield.library.upenn.edu/histcomp/sciento_all_citing/index-lcs-4.html
http://www.garfield.library.upenn.edu/histcomp/sciento_all_citing/index-lcs-4.html

Description ... cognitive perspective in information science ... SCIENCE 63(5):261-269.
NARANAN S; BALASUBRAHMANYAN VK. INFORMATION THEORETIC MODELS IN
STATISTICAL LINGUISTICS ... AND COMPUTER SCIENCES 37(5 ...

Compass Results

Search terms: computational linguistics discipline computer science aspects human language cognitive sciences intelligence aiming models applied

Date Tue Apr 25 20:04:55 EDT 2006

CL Intro Text

http://www.coli.uni-saarland.de/~hansu/what_is_cl.html

Description WHAT IS COMPUTATIONALLINGUISTICS? ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... linguistics has applied ...

Score: 0.8237344465051767

NLP Reading

<http://www.infm.ulst.ac.uk/~MAEVE/nlp/reading.html>

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Score: 0.7806636948950251

Computers and the English language

<http://elex.amu.edu.pl/~przemka/Lectures/Lecture02pk.htm>

Description ... computer-human interaction ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... computational models of ...

Score: 0.7671556262282142

Newsgroup FAQ's: comp.ai.nat-lang/Natural Language Processing FAQ

http://www.newsville.com/cgi-bin/getfaq?file=comp.ai.nat-lang/Natural_Language_Processing_FAQ

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Score: 0.7293899381619214

Newsgroup FAQ's: comp.ai/Natural Language Processing FAQ

http://www.newsville.com/cgi-bin/getfaq?file=comp.ai/Natural_Language_Processing_FAQ

Description ... discipline between linguistics and. computer science which is concerned with the computational aspects of the. human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Score: 0.729389938161921

Natural Language Processing FAQ

<http://omicron.felk.cvut.cz/FAQ/articles/a1248.html>

Description ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Score: 0.7285778153626946

Natural Language Processing FAQ

http://icl.pku.edu.cn/icl_members/member/zwd/NLP/OVERVIEW/NLP-FAQ.htm

Description Natural Language Processing FAQ. From: radev@tune.cs.columbia.edu (Dragomir R. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... CL: applied and ...

Score: 0.7283396287194188

Formal Methods

<http://nora.hd.uib.no/AcoHum/fm/fm-chapter-final.html>

Description ... with Applied Computing ... Cognitive and Computing Sciences, offers humanities related subjects including Linguistics (including computer linguistics), philosophy, and Computer Science ... of human language ...

Score: 0.700889118368703

Linguist List - Web Resource Listings

<http://linglist.org/sp/Projects.html>

Description Fund Drive 2006 has set sail! Projects & Research Sites ... the Cognitive Science Building on ... Computational Linguistics: Department of Computer Science, Technion, Israel Institute of Technology. Lakota Language ... computational models of human ...

Score: 0.6887352623016749

Linguist List - Web Resource Listings

<http://www.ling.ed.ac.uk/linguist/sp/Projects.html>

Description ... the Cognitive Science Building on ... Computational Linguistics: Department of Computer Science, Technion, Israel Institute of Technology. Lakota Language ... computational models of human ...

Score: 0.6887352623016749

Science - Social Sciences - Linguistics - Computational Linguistics - Newsletter - News - Reviews - Education - ...

http://www.banner-net.com/Science/Social_Sciences/Linguistics/Computational_Linguistics

Description ... done aiming for ... of human communication ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...

Score: 0.6776001010479363

Introductory

<http://www.msu.edu/course/lin/475/notes/hale-notes002.html>

Description 1.1 What is Computational Linguistics? ... psychological models of ... modeling human linguistic ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...
Score: 0.652736181829402

Présentation de la journée.
<http://www.olats.org/space/seti/2003/space13Avr2003.html>

Description Présentation. On March 23-24, 2003, the second in a series of international workshops on interstellar message design will be held in Paris. ... Human-Computer Interaction (HCI). He joined the faculty at the University of Birmingham in 1986 to research and teach Cognitive Science (linguistics ... computational models of ...
Score: 0.6431365219961638

News Indexed by Topic - COGNITIVE SCIENCE ARCHIVE
<https://www.aai.org/AITopics/newsttopics/cognitive1.html>

Description AI Topics provides basic, understandable information and helpful resources concerning artificial intelligence, with an emphasis on material available online. ... computational sciences is the understanding of the mechanisms of the human ... new discipline of cognitive science, the ... computer." >>> Cognitive Science, Nature of Intelligence ...
Score: 0.6305096020209757

Littera:
<http://littera.deusto.es/prof/abaitua/hlt/hlt0304/Computational%20Linguistics>

Description Computational Linguistics is a subfield of Linguistics in which logical modelling of natural language from an computational perspective is central. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ... linguistics has applied ...
Score: 0.5872141472147729

MITECS: Culture, Cognition, and Evolution
<http://www.cs.buffalo.edu/pub/WWW/faculty/rapaport/575/FO1/mycultureintro.html>

Description ... cognitive sciences focuses ... more human-like ... human mind as essentially a general intelligence. Arguments and evidence from evolutionary theory, developmental psychology, linguistics, and one approach in cognitive ...
Score: 0.5865220294178932

Year 2006 News Archive: February
<http://www.aai.org/aitopics/html/archvG2.html>

Description AI Topics provides basic, understandable information and helpful resources concerning artificial intelligence, with an emphasis on material available online. ... Models Computer on ... human terms. Moss said he'd also like to initiate collaborations with the Computer Science and Artificial Intelligence ...
Score: 0.5857171322084863

IBRI Research Report #31 - BEYOND THE SHADOW OF A DOUBT: Logical Deduction and the Reasoning Process

<http://www.ibri.org/RRs/RR031/31shadow.html>

Description ... The science of artificial intelligence as applied to image ... a human intelligence in all measurable aspects, then for ... and linguistics, and other branches of cognitive science'). What ...

Score: 0.580958274019837

EpistemeLinks: For Philosophy Resources on the Internet

<http://www.epistemelinks.com/Main/Journals.aspx?Publisher=Kluwer>

Description EpistemeLinks is a comprehensive resource for philosophy on the Internet, providing thousands of links categorized by philosopher, topic, and resource type. ... Intelligence, Cognitive Science and ... science in general and natural and human sciences ... aspects of artificial intelligence concerned with language such as computational linguistics ...

Score: 0.5730286488889133

My Research Area

<http://www.cfilt.iitb.ac.in/~debasri/research.html>

Description Research Interests. My Research Area is Natural Language Processing. This also known as Computational Linguistics. ... discipline between linguistics and computer science which is concerned with the computational aspects of the human language faculty. It belongs to the cognitive sciences ...

Score: 0.5620622608067227

ACT | Departments | Philosophy and Social Sciences | Course Descriptions

http://www.anatolia.edu.gr/act/academics/departments/philsoc/course_descriptions.shtml

Description Course Descriptions. Philosophy. Philosophy 101: Introduction to Philosophy and Critical Reasoning. The primary aim of this course is to train students in the skills required for critical analysis of discourse. ... Cognitive Science. Researchers in psychology, linguistics, computer ... aspects of human intelligence ... sciences and one of anthropology's distinctive characteristics as a discipline ...

Score: 0.5592484267794919

Computer Laboratory - The Blue Book - University Teaching Officers

<http://www.cl.cam.ac.uk/DeptInfo/Overview/node4.html>

Description ... cognitive science and ... computational and theoretical linguistics and automated speech and language processing. He directs and is heavily involved in the teaching of the MPhil in Computer ...

Score: 0.5540098309066683

Psychology - Wikipedia, the free encyclopedia

<http://en.wikipedia.org/wiki/Psychology>

Description ... models of human ... computer science and neuroscience) in the quest to understand the mind, the umbrella discipline of cognitive ... models from cognitive psychology are widely applied ...

Score: 0.5321680681316237

The KLI Theory Lab - Ethology, Sociobiology
<http://www.kli.ac.at/theorylab/Areas/ETSB.html>

Description KLI Theory Lab — Area Introductions. ETSB. ETHOLOGY AND. SOCIOBIOLOGY. Even though our bodies and behaviors share many common attributes, it's far more fruitful to consider not one human nature but many. ... human behavior, aiming to ... cognitive sciences, comparative ... cognitive science, including artificial intelligence, human information processing, linguistics, mathematical models ...
Score: 0.5313236100969475

Machine Learning List: Vol. 9, No. 16
<http://www.ics.uci.edu/~mlearn/MLlist/v9/16.html>

Description ... applied methods from statistics, computer science, artificial intelligence ... complementary aspects of ... Intelligence and Cognitive Science, including Cognitive Psychology, Computational Linguistics ...
Score: 0.5291665298810726

THE CHALLENGE OF SOCIOCYBERNETICS
http://uwacadweb.uwyo.edu/RED_FEATHER/chaos/oo6challenges.html

Description ... have applied it to the social sciences: Deutsch2 ... mental" models of their environment that act like computer subroutines ... Embodied Mind - Cognitive Science and Human Experience (3rd ...
Score: 0.521733777987543

http://www.garfield.library.upenn.edu/histcomp/sciento_all_citing/index-lcs-4.html
http://www.garfield.library.upenn.edu/histcomp/sciento_all_citing/index-lcs-4.html

Description ... cognitive perspective in information science ... SCIENCE 63(5):261-269. NARANAN S; BALASUBRAHMANYAN VK. INFORMATION THEORETIC MODELS IN STATISTICAL LINGUISTICS ... AND COMPUTER SCIENCES 37(5 ...
Score: 0.4877955199622039

Psychology - Wiki
<http://www.yikipedia.org/mediawiki/index.php/Psychology>

Description Psychology. From Wiki. Psychology. Areas. Lists ... studying human psychology ... Cognitive science again considers the "mind" as a subject for investigation, using the tools of evolutionary psychology, linguistics, computer ... computational models to ...
Score: 0.47839083104282976

Psychology/rewrite - Wikipedia, the free encyclopedia
<http://en.wikipedia.org/wiki/Psychology/rewrite>

Description Psychology/rewrite From Wikipedia, the free encyclopedia This article is in the process of being merged into Psychology, and may be outdated. ... human psychology, known collectively as cognitive ... linguistics, computer science ... applied universally to psychology, as some areas within the discipline, such as cognitive science ...
Score: 0.47807901499363187

TopNavBar

http://www.olats.org/space/seti/2003/mono_index.html

Description ... Extraterrestrial Intelligence (SETI ... Human-Computer Interaction (HCI). He joined the faculty at the University of Birmingham in 1986 to research and teach Cognitive Science (linguistics ... computational models of ...

Score: 0.0

Acknowledgments:

I would like to thank my supervisor Dr. Sabine Bergler, for her suggestions, encouragement and critical eye in reading, listening and responding to my work.

References:

1. G. Salton, A. Wong, and C. S. Yang (1975), A Vector Space Model for Automatic Indexing, *Communications of the ACM*, vol. 18, nr. 11, pages 613–620
2. C. Manning and H Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999
3. Y. Matsuo and M. Ishizuka. Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information, Int'l Journal on Artificial Intelligence Tools, Vol.13, No.1, pp.157-169, 2004
4. Porter, M. F. "An Algorithm for Suffix Stripping." Program 14, 1980, 130--137.
5. C. J. van Rijsbergen. Information Retrieval. Butterworths, London, 1979