Assisted Query Formulation using Normalised Word Vector and Dynamic Ontological Filtering

Heinz Dreher, Robert Williams

Curtin University of Technology, GPO Box U1987
Perth, Western Australia 6845
{Heinz.Dreher, Bob.Williams}@cbs.curtin.edu.au

Abstract. Information seekers using the usual search techniques and engines are delighted by the sheer power of the technology at their command – speed, quantity. Upon closer inspection of the results, and reflection upon the next stages of the information seeking knowledge work, users are typically overwhelmed, and frustrated. We propose a partial solution by focusing on the query formulation aspect of the information seeking problem. First we introduce our version of a semantic analysis algorithm, named Normalised Word Vector, and explain its application in assisted query formulation. Secondly we introduce our ideas of supporting query refinement via Dynamic Ontological Filtering.

1 Introduction

Information and Communications Technologies (ICT) pervade our society in which a growing proportion of the work is mental work (typically referred to as knowledge work) as opposed to physical work. The record of empowerment of physical workers through technology clearly shows the enormous benefit in terms of increased production, efficient and effective utilisation of resources, and greater safety for workers. Knowledge workers should expect to see similar gains in capacity, productivity, and in the quality of outputs, but progress in the empowerment of the knowledge worker needs a boost so that this promise and expectation can be realised. It is not so much a case of lack of vision:

> Neither the naked hand nor the understanding left to itself can effect much. It is by instruments and helps that the work is done, which are as much wanted for the understanding as for the hand. And as the instruments for the hand either give motion or guide it, so the instruments of the mind supply either suggestions for the understanding or cautions. [1].

More recently, by some three and a quarter centuries, Vannevar Bush [2] shared with us his vision of how research workers could be empowered with his MEMEX device. But despite the vision, and the tremendous advances in ICT, and their now pervasive nature, the knowledge worker is left languishing by and large. For example, whilst a literature search can now be conducted in a matter of days if not hours where only two decades ago it took weeks if not months, the researcher is confronted with millions upon millions of ‘hits’.

Present methods of refinement of the result of a query or search are inadequate – there is far too much material which is potentially relevant. Additionally, users, don't really know what is relevant until some way through the discovery, learning, or knowledge acquisition process. Help is needed with the Query-formulation → Find → Re-formulation phases of knowledge work.

In this article we consider the query-formulation aspect of the general problem. There are two contributions we propose to integrate into search and find processes. Firstly, an adaptation of a semantic analysis algorithm named Normalised Word Vector (NWV) developed by Williams [7] for the MarkIT (www.essaygrading.com) Automated Essay Grading project [8] with the aim of accepting natural language query expressions. Secondly, we propose a dynamic ontological filter to be applied to the query in order to maximise search relevance. Categorizing the results returned by search engines and presenting the categories to the user through a special browser endowed with an ontology navigation scheme is expected to contribute to a refinement of the search query and hence improve relevance and thus information quality as required by the user.

2 Assisted Query Formulation – empowering P to refine Q

Imagine a human user P is interested in researching Yoga, and enters this as a search-term (Q) into Google which delivers circa 39 million results in one tenth of a second. This is impressive until P begins to use the returned results. P quickly determines that some strategy is needed to reduce the quantity and increase the relevance of the results, but to accomplish this P will need to refine Q, and proceeds using the traditional methods such as including Boolean operations, appending adjectives, enclosing search strings within quotes, and so on. P's focus has now been redirected from the concept "Yoga", as originally envisaged, to the science of search-term and query formulation. Actually, we all know that P did not have just "yoga' in mind, but some idiosyncratic aspect/s thereof. Surely we can better support P in the information seeking task pertaining to Yoga.

Consider the 7-Step process in Table 1, which we have termed Query-formulation → Find → Re-formulation (QFR).

Table 1. Query-formulation → Find → Re-formulation (QFR)

Step 1)	Person **P** has idea → constructs keywords or some text to explain the concept → call it **Q**.
Step 2)	Via a special browser, **Q** is 'acquired' by the NWV technology which makes computations based on some 'reference data' set. This is typically a thesaurus or alternative corpus obtained by some search-categorization process or by reference to a seed ontology - the Open Directory Project (www.dmoz.com) would be a suitable starting point.
Step 3)	A set of *Normalised Concepts* are returned in the context of the 'reference data' set and categorized by the reference to the current ontological view.

Step 4)	The special browser facilitates **P** to adjust (augment, amend, re-arrange, re-categorize, delete) **Q**, we now have **Q**i (i ranges from 1 for the first iteration to integer values such as 3 or 4, perhaps 7 at most).
Step 5)	Re-iterate through Steps 2) 3) and 4) until **P** is satisfied that **Q**n represents the true idea **P** had in mind for the search.
Step 6)	Submit **Q**n into ontological filter/disambiguation system to match the query/ontology/target data repository for the search.
Step 7)	Present results to **P** with options of re-iterating Step 6) after **P** refines/adjusts and/or repeats Step 5).

Continuing our Yoga example from above, at Step 1) we have P composing a natural language query into the special browser.

Q = I would like to take an English language based Yoga-teacher course as soon as possible

(Google returns 900 odd results in about half a second, and if Q is enclosed in quotes, Google returns no matching documents, also quite speedily).

The NWV technology computes Normalised Concepts based on Q, which for example may be Language, Lifestyle, Time, Education. These constitute the 'core concepts' contained in Q - comprising Step 2). Varied forms of Q as expressed by various P would all yield the same Normalised Concepts thereby already greatly simplifying and focusing the subsequent document match and retrieval. At this point one may mention that the 'core concepts' could be readily translated into arbitrarily many natural languages, and depending on the respective thesaurus (or alternative) structures, similarly high quality results could be expected for those alternative language searches. Further, with such a simplified query form, it is possible to imagine a new FAQ database which efficiently supports Normalised Concept searching.

Step 3) of the QFR in Table 1 also categorizes the Normalised Concepts according to the currently activated ontological filter. At Step 4) our user P interacts with the content (Normalised Concepts) and with the structure (ontological view) of the idea behind the search query Q.

P repeats some steps until a much finer and accurate representation of the original Q is created. For example, P realised that Yoga courses everywhere in the world were not of interest, but rather, Yoga courses in Milano, Italy, or even more specifically in either San Giovanni or Corsico, respectively, on the north east and south west fringes of the Milano metropolis.

Now at Step 6), a truly proper representation of P's Q exists and is expected to result in a high quality outcome at Step 7).

One might say that the technology helps watch over the P's and Q's. And this is properly as it should be, as compared with the user confronted with millions of documents returned in practically millionths of seconds – here the user is being supported by the technology rather than the technology dictating how/what the user should or can do. We would argue that our "Assisted Query Formulation using Normalised Word Vector and Dynamic Ontological Filtering" is truly empowering.

In the remainder of the paper we describe the two main components of our system: the Normalised Word Vector technology and the Dynamic Ontological Filtering, and give some examples of the technology at work.

3 The Normalised Word Vector Technology

NWV consists of two parts. First there is the thesaurus based vector representation which is explained via an example found in Tables 2, 3, and 4, and Fig. 1. Once the vector representations have been built, a document similarity measure may be calculated using geometrical constructs.

3.1 Thesaurus based Vector Representation to Normalise Words in Documents

Vector algebra techniques are used to represent similarities in content between documents or natural language free text expressions. A thesaurus [5] is used to build this vector representation by 'normalising' the words in the documents by reducing all words to a thesaurus root word appropriate to the encompassing concept. The vector representation is then constructed from the enumeration of these concepts. In our work on automated essay grading [3], [8], we have shown the approach to perform satisfactorily; that is, as well as one would expect from humans. The following start-of-sentence fragments from successive sentences in three separate Document Texts are used to explain how this is accomplished.

Table 2. Document number and text

Document Number	Document Text
(1)	The tall girl… A large female
(2)	A major girl… A happy boy
(3)	The large girl… Some major holiday

Suppose a thesaurus exists with the following root Concept Numbers and Words:

Table 3. Concept number and words

Concept Number	Words
1.	the, a
2.	tall, large, major
3.	girl, female
4.	little
5.	happy
6.	boy
7.	some
8.	holiday

Three dimensional vector representations of the above document fragments on the first three Concept Numbers (1-3) can be constructed by counting the number of times a word belonging to that Concept Number appears in the document fragments. These vectors are given in Table 4.

Table 4. Document number and matching concepts

Document Number	**Vector on first 3 concepts**	**Explanation**
(1)	[2, 2, 2]	[The, a; tall, large; girl, female]
(2)	[2, 1, 1]	[A, a; major; girl]
(3)	[1, 2, 1]	[The; large, major; girl]

Fig. 1 presents these 3-dimensional vectors pictorially. Document Number (1) is the dotted line emerging from the origin at the bottom left and stretching to the upper right of the graph. Document Numbers (2) and (3) are to the left and right, respectively, of (1). Concepts 1, 2, and 3, are shown as orthogonal axes as a reference frame. The angles, Theta, are a measure of the separation between documents - Theta1 between (2) and (3) and Theta2 between (1) and (3).

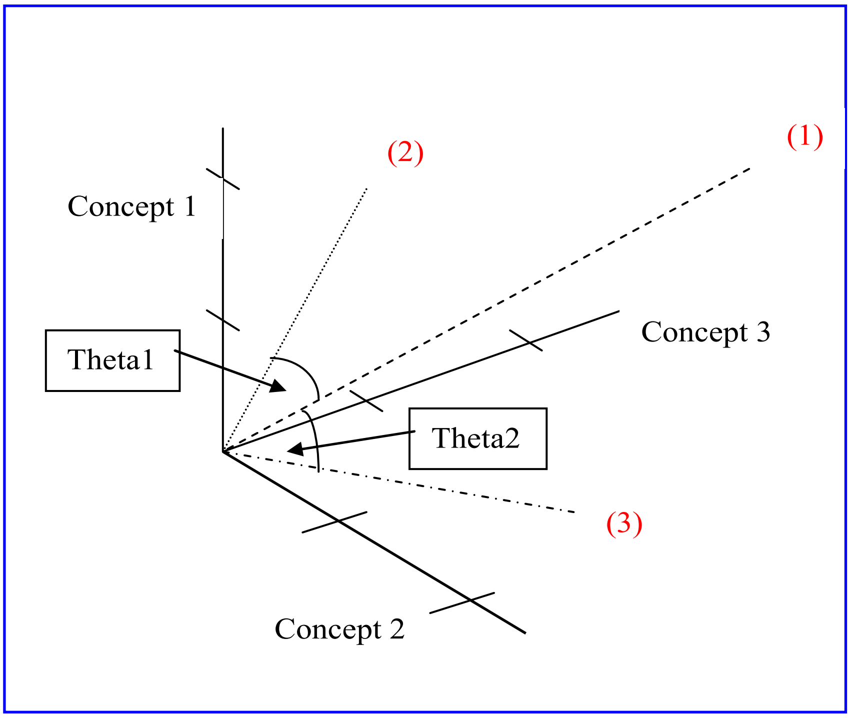


Fig. 1. Vector representation (broken lines) of documents

3.2 Document Similarity - Computing the Variable CosTheta

The closeness in terms of the semantics between documents (1) and (2), and (1) and (3) can be determined by looking at the closeness of their corresponding vectors. The angle between the vectors varies according to how 'close' the vectors are. A small angle indicates that the documents contain similar content; a large angle indicates that they do not have much common content. This measure of closeness can be quantified by looking at the cosines of Theta1 and Theta2. If documents (1) and (2) were identical, their vectors would be identical, and would be collinear, resulting in a cosine value of 1. If on the other hand, they were completely different, and therefore orthogonal, their cosines would be 0.

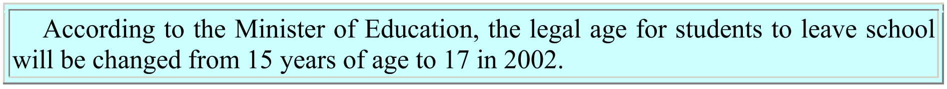
The variable named CosTheta used in the NWV algorithm is this cosine computed for the document (semantic content) being evaluated.

The Macquarie Thesaurus from The Macquarie Library Pty Ltd [5] is used to derive the normalised concepts. There are 812 concepts in this thesaurus, and all words in the documents are reduced to the appropriate within-context root concept. The vectors are constructed in this 812 dimensional space, and the vector theory carries over to these dimensions in exactly the same way – it is of course somewhat challenging to visualize the vectors in this 812D-hyperspace, as they cannot be represented graphically.

4 Applying NWV Technology to Query-formulation

The reader can see that the scenario from an earlier section is readily supported by NWV technology. The documents referred to in Fig. 1 become the successive versions of **Q**, the query-formulation being constructed by the researcher **P**.

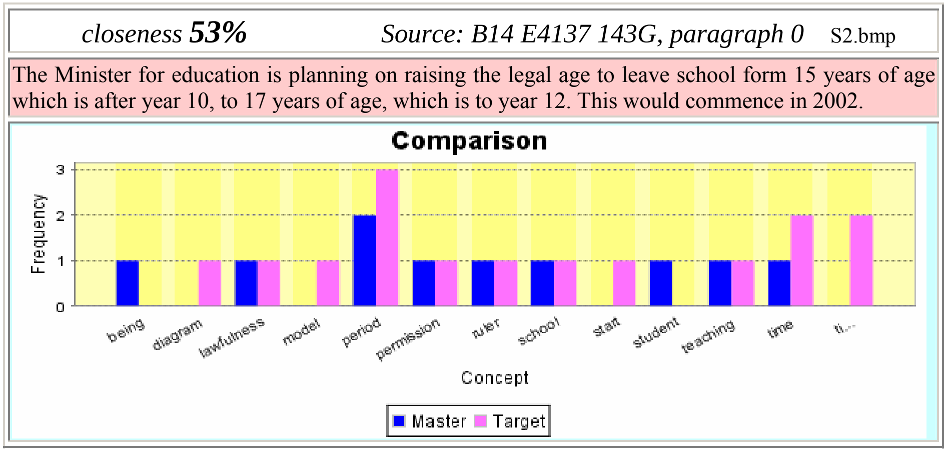
Next we consider a test of the NWV system of query-formulation based on expressions taken from essays written by year 10 Western Australian high school students on the topic of "The School Leaving Age". For example, suppose the user is interested in finding documents relating to the following content:



According to the Minister of Education, the legal age for students to leave school will be changed from 15 years of age to 17 in 2002.

Fig. 2. "The School Leaving Age" query formulation

The NWV system returned the following alternate natural language or textual query formulations by other 'researchers' together with a closeness representation. Four cases are presented in increasing order of 'semantic closeness' – 53%, 54%, 66% and 77%. In the figures below (see Legend), the "Master" corresponds to the Concepts derived from the text in Fig. 2 (blue, dark), and the "Target" corresponds to the text found in Figs. 3, 4, 5, and 6 (magenta or pink).

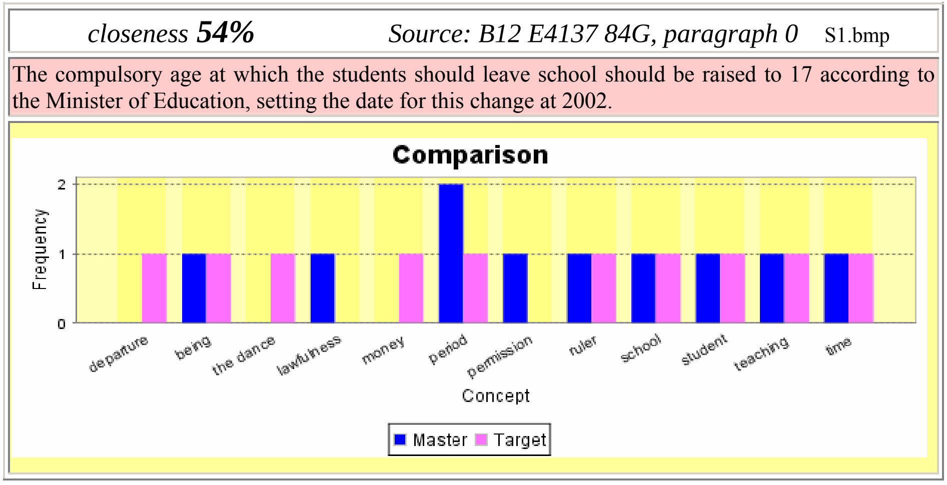


closeness ***53%*** *Source: B14 E4137 143G, paragraph 0* S2.bmp

The Minister for education is planning on raising the legal age to leave school form 15 years of age which is after year 10, to 17 years of age, which is to year 12. This would commence in 2002.

Fig. 3. 53% semantic proximity

A comparison of the ‘Master’ dark blue coloured bar (left pair member) with the ‘Target’ light or magenta coloured bar (right pair member) affords a visual comparison of the semantic proximity by Concept. In the case of the Target text given in Fig. 3 above, which has 53% semantic commonality with the Master text given in Fig. 2, we can observe the numerous concept mismatches (“being”, “diagram”, “model”, “start”, “student”, and “ti…”). A similar mismatch, in relative terms, is observed in Fig. 4 below (note the change in vertical axis scale).

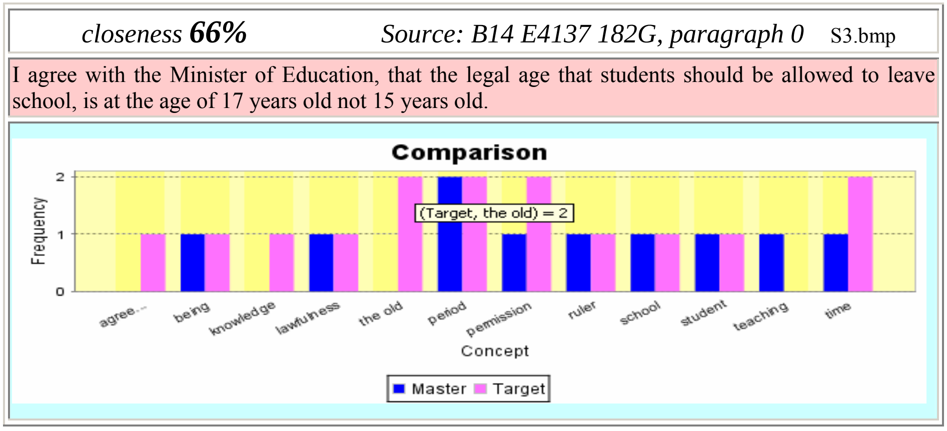


closeness ***54%*** *Source: B12 E4137 84G, paragraph 0* S1.bmp

The compulsory age at which the students should leave school should be raised to 17 according to the Minister of Education, setting the date for this change at 2002.

Fig. 4. 54% semantic proximity

The NWV technology is doing its job of representing the free text found in documents and documents fragments according to the conceptual structure found in the thesaurus. In this case we are using the Macquarie Thesaurus [5] but any structured corpus can be used. A visual representation is useful to the researcher **P** because the change in **Q** over the iterations can be modeled and interacted with to produce a superior future iterative version of **Q**.



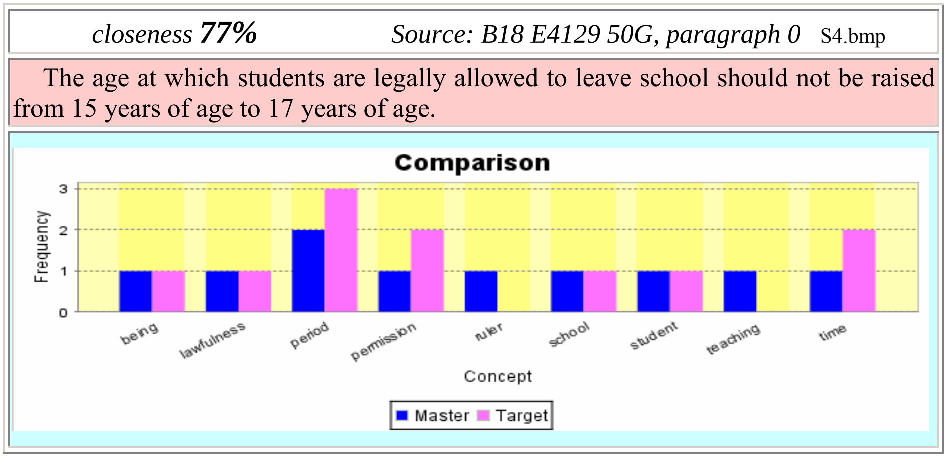
closeness ***66%*** *Source: B14 E4137 182G, paragraph 0* S3.bmp

I agree with the Minister of Education, that the legal age that students should be allowed to leave school, is at the age of 17 years old not 15 years old.

Fig. 5. 66% semantic proximity

In the last of the four cases of alternate query formulations, NWV returns a 77% degree of semantic closeness. Observe that there are only two mismatching concepts ("ruler", and "teaching") and that the frequency corresponds rather well also.

We are satisfied that NWV technology is supporting our purpose. Readers should take a moment or two to review the textual query formulations as given in these four figures and compare the meanings with that derived from the text in Fig. 2.



closeness ***77%*** *Source: B18 E4129 50G, paragraph 0* S4.bmp

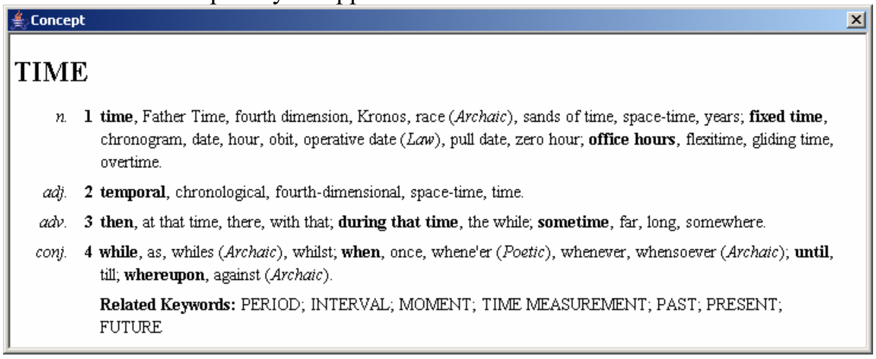
The age at which students are legally allowed to leave school should not be raised from 15 years of age to 17 years of age.

Fig. 6. 77% semantic proximity

The sequence depicted in Figs. 3 through 6 can be seen as a refinement of a query formulation with respect to a 'Master'. Of course this sequence has been artificially composed to emphasize the power of the NWV technology. Note how the Master and Target concepts coincide as one progresses through the 53% match in Fig. 3 to the 77% match in Fig. 6.

Naturally, our explanation is incomplete without some further treatment of 'Concepts'.

We refer the reader to the following five Concept windows (Fig. 7 through Fig. 11) generated from the www.essaygrading.com site. They are the entries as produced by the Macquarie Thesaurus [5], the Concept Name in the window matching the Concept labels on the horizontal axes of the bar chart representations above. Inspection of the Concept window contents will permit an understanding of the complexity of the task on the one hand and hopefully an appreciation of the semantics in addition.



TIME

n. **1** **time**, Father Time, fourth dimension, Kronos, race (*Archaic*), sands of time, space-time, years; **fixed time**, chronogram, date, hour, obit, operative date (*Law*), pull date, zero hour; **office hours**, flexitime, gliding time, overtime.

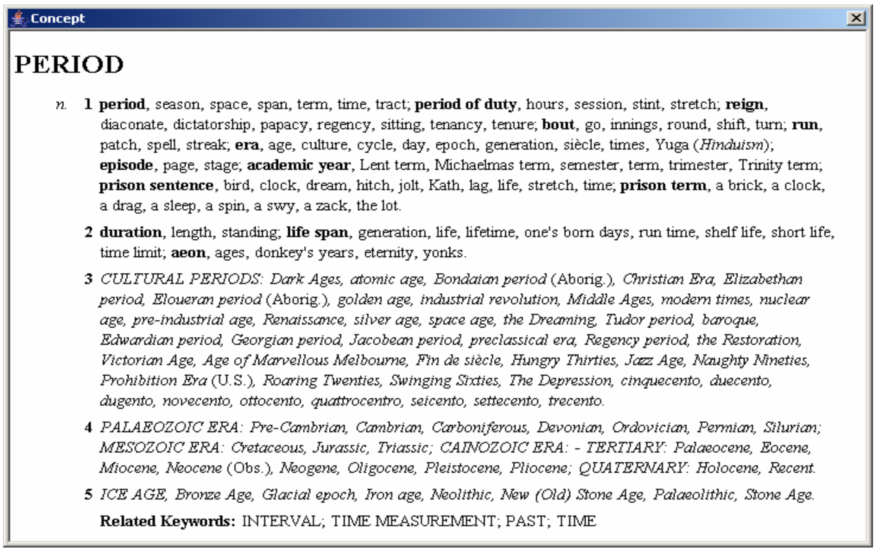
adj. **2** **temporal**, chronological, fourth-dimensional, space-time, time.

adv. **3** **then**, at that time, there, with that; **during that time**, the while; **sometime**, far, long, somewhere.

conj. **4** **while**, as, whiles (*Archaic*), whilst; **when**, once, whene'er (*Poetic*), whenever, whensoever (*Archaic*); **until**, till; **whereupon**, against (*Archaic*).

Related Keywords: PERIOD; INTERVAL; MOMENT; TIME MEASUREMENT; PAST; PRESENT; FUTURE

Fig. 7. The concept TIME



PERIOD

n. **1** **period**, season, space, span, term, time, tract; **period of duty**, hours, session, stint, stretch; **reign**, diaconate, dictatorship, papacy, regency, sitting, tenancy, tenure; **bout**, go, innings, round, shift, turn; **run**, patch, spell, streak; **era**, age, culture, cycle, day, epoch, generation, siècle, times, Yuga (*Hinduism*); **episode**, page, stage; **academic year**, Lent term, Michaelmas term, semester, term, trimester, Trinity term; **prison sentence**, bird, clock, dream, hitch, jolt, Kath, lag, life, stretch, time; **prison term**, a brick, a clock, a drag, a sleep, a spin, a swy, a zack, the lot.

2 **duration**, length, standing; **life span**, generation, life, lifetime, one's born days, run time, shelf life, short life, time limit; **aeon**, ages, donkey's years, eternity, yonks.

3 *CULTURAL PERIODS: Dark Ages, atomic age, Bondaian period* (Aborig.), *Christian Era, Elizabethan period, Eloueran period* (Aborig.), *golden age, industrial revolution, Middle Ages, modern times, nuclear age, pre-industrial age, Renaissance, silver age, space age, the Dreaming, Tudor period, baroque, Edwardian period, Georgian period, Jacobean period, preclassical era, Regency period, the Restoration, Victorian Age, Age of Marvellous Melbourne, Fin de siècle, Hungry Thirties, Jazz Age, Naughty Nineties, Prohibition Era* (U.S.), *Roaring Twenties, Swinging Sixties, The Depression, cinquecento, duecento, dugento, novecento, ottocento, quattrocentro, seicento, settecento, trecento.*

4 *PALAEOZOIC ERA: Pre-Cambrian, Cambrian, Carboniferous, Devonian, Ordovician, Permian, Silurian; MESOZOIC ERA: Cretaceous, Jurassic, Triassic; CAINOZOIC ERA: - TERTIARY: Palaeocene, Eocene, Miocene, Neocene* (Obs.), *Neogene, Oligocene, Pleistocene, Pliocene; QUATERNARY: Holocene, Recent.*

5 *ICE AGE, Bronze Age, Glacial epoch, Iron age, Neolithic, New (Old) Stone Age, Palaeolithic, Stone Age.*

Related Keywords: INTERVAL; TIME MEASUREMENT; PAST; TIME

Fig. 8. The concept PERIOD

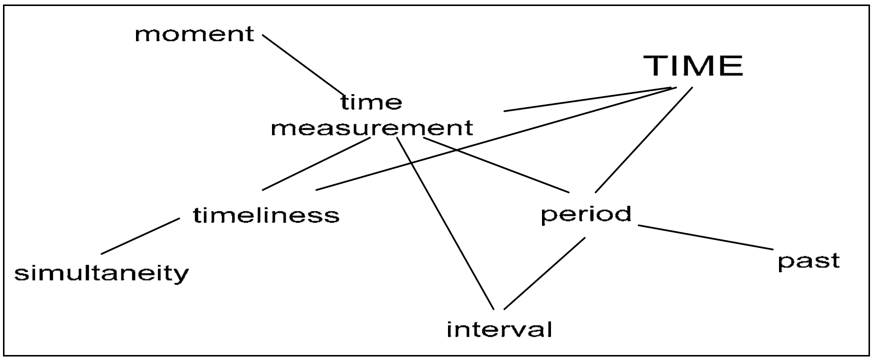
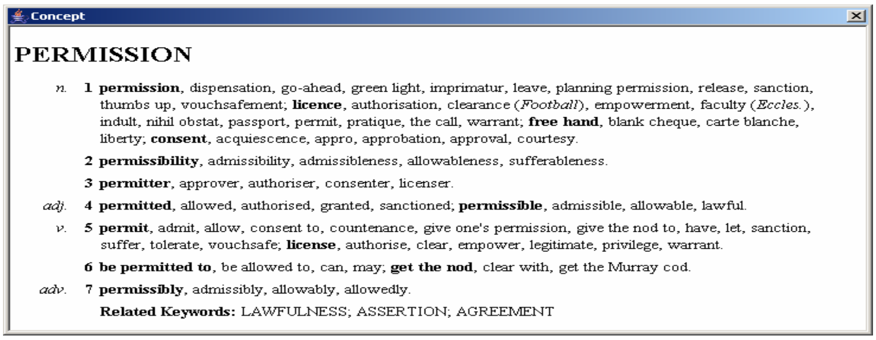


Fig. 9. The concept mapping for TIME and its relatives



PERMISSION

n. **1** **permission**, dispensation, go-ahead, green light, imprimatur, leave, planning permission, release, sanction, thumbs up, vouchsafement; **licence**, authorisation, clearance (*Football*), empowerment, faculty (*Eccles.*), indult, nihil obstat, passport, permit, pratique, the call, warrant; **free hand**, blank cheque, carte blanche, liberty; **consent**, acquiescence, appro, approbation, approval, courtesy.

2 **permissibility**, admissibility, admissibleness, allowableness, sufferableness.

3 **permitter**, approver, authoriser, consenter, licenser.

adj. **4** **permitted**, allowed, authorised, granted, sanctioned; **permissible**, admissible, allowable, lawful.

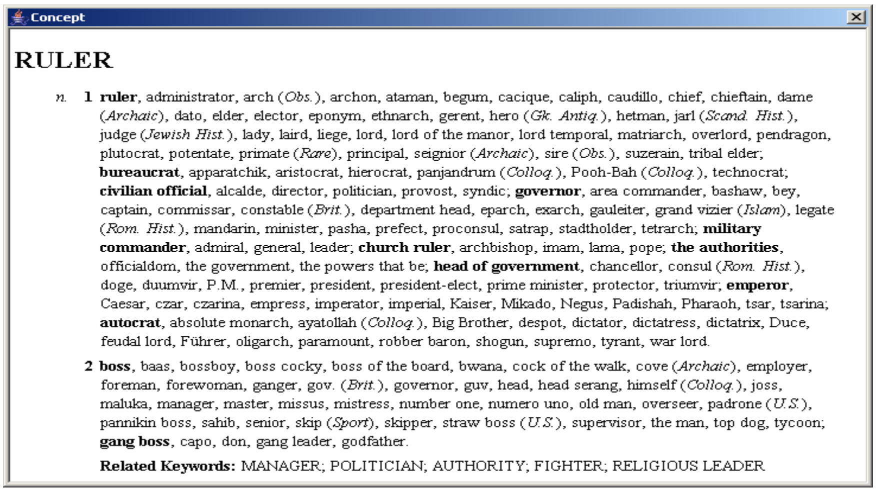
v. **5** **permit**, admit, allow, consent to, countenance, give one's permission, give the nod to, have, let, sanction, suffer, tolerate, vouchsafe; **license**, authorise, clear, empower, legitimate, privilege, warrant.

6 **be permitted to**, be allowed to, can, may; **get the nod**, clear with, get the Murray cod.

adv. **7** **permissibly**, admissibly, allowably, allowedly.

Related Keywords: LAWFULNESS; ASSERTION; AGREEMENT

Fig. 10. The concept PERMISSION



RULER

n. **1** **ruler**, administrator, arch (*Obs.*), archon, ataman, begum, cacique, caliph, caudillo, chief, chieftain, dame (*Archaic*), dato, elder, elector, eponym, ethnarch, gerent, hero (*Gk. Antiq.*), hetman, jarl (*Scand. Hist.*), judge (*Jewish Hist.*), lady, laird, liege, lord, lord of the manor, lord temporal, matriarch, overlord, pendragon, plutocrat, potentate, primate (*Rare*), principal, seignior (*Archaic*), sire (*Obs.*), suzerain, tribal elder; **bureaucrat**, apparatchik, aristocrat, hierocrat, panjandrum (*Colloq.*), Pooh-Bah (*Colloq.*), technocrat; **civilian official**, alcalde, director, politician, provost, syndic; **governor**, area commander, bashaw, bey, captain, commissar, constable (*Brit.*), department head, eparch, exarch, gauleiter, grand vizier (*Islam*), legate (*Rom. Hist.*), mandarin, minister, pasha, prefect, proconsul, satrap, stadtholder, tetrarch; **military commander**, admiral, general, leader; **church ruler**, archbishop, imam, lama, pope; **the authorities**, officialdom, the government, the powers that be; **head of government**, chancellor, consul (*Rom. Hist.*), doge, duumvir, P.M., premier, president, president-elect, prime minister, protector, triumvir; **emperor**, Caesar, czar, czarina, empress, imperator, imperial, Kaiser, Mikado, Negus, Padishah, Pharaoh, tsar, tsarina; **autocrat**, absolute monarch, ayatollah (*Colloq.*), Big Brother, despot, dictator, dictatress, dictatrix, Duce, feudal lord, Führer, oligarch, paramount, robber baron, shogun, supremo, tyrant, war lord.

2 **boss**, baas, bossboy, boss cocky, boss of the board, bwana, cock of the walk, cove (*Archaic*), employer, foreman, forewoman, ganger, gov. (*Brit.*), governor, guv, head, head serang, himself (*Colloq.*), joss, maluka, manager, master, missus, mistress, number one, numero uno, old man, overseer, padrone (*U.S.*), pannikin boss, sahib, senior, skip (*Sport*), skipper, straw boss (*U.S.*), supervisor, the man, top dog, tycoon; **gang boss**, capo, don, gang leader, godfather.

Related Keywords: MANAGER; POLITICIAN; AUTHORITY; FIGHTER; RELIGIOUS LEADER

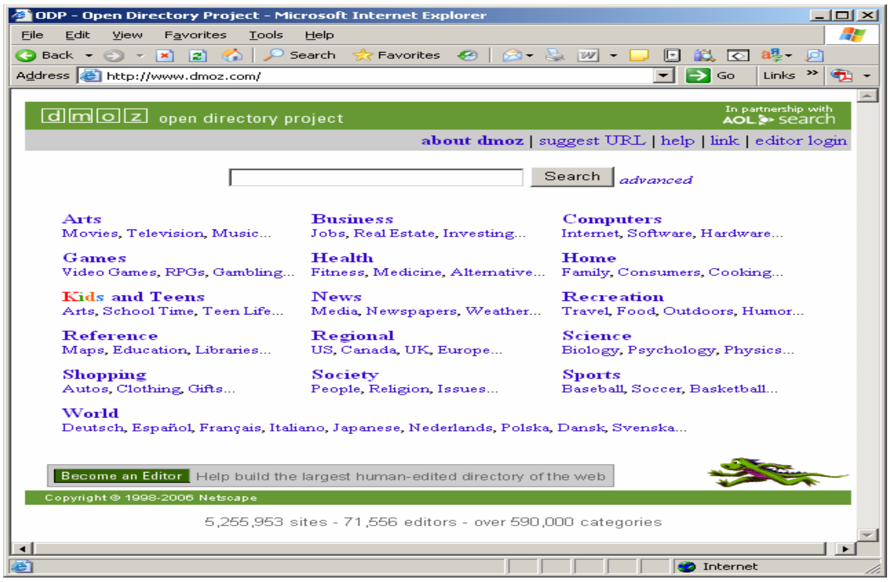
Fig. 11. The concept RULER

5 Dynamic Ontological Filtering

Our research in this second aspect of the work is still very much in progress, however the idea we envisage is as follows. All concepts which **P** wishes to include in the search as represented in **Q** will be found in a subset of the world's knowledge. Which subset and indeed which portion of a subset will be determined with the help of an ontological filter. Obviously one must have a target body of knowledge to search and then one would ideally have some knowledge structuring device such as a purpose built ontology, however if this does not exist we can generate one dynamically based on a starting point or seed ontology.

Ide & Veronis [4] explain "In general terms, *word sense disambiguation* (WSD) involves the association of a given word in a text or discourse with a definition or meaning (*sense*) which is distinguishable from other meanings potentially attributable to that word". Ontology based Query disambiguation or filtering will refine and re-organize search results according to their similarity with the thematic content of appropriate categories of that ontology – the ODP in our case [6].

The focus of the research is to ontologically disambiguate search query **Q** by categorizing search results returned by search engines such as Google or Yahoo. We propose to use the most comprehensive human-edited directory of the Web as represented in The Open Directory Project (www.dmoz.com), effectively as a starting point. The Open Directory Project's 16 level top level hierarchy is shown in Fig. 12.


ODP - Open Directory Project - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.dmoz.com/
Go Links
dmoz open directory project
In partnership with AOL search
about dmoz | suggest URL | help | link | editor login
Search advanced
Become an Editor
Internet

Arts
Movies, Television, Music...

Business
Jobs, Real Estate, Investing...

Computers
Internet, Software, Hardware...

Games
Video Games, RPGs, Gambling...

Health
Fitness, Medicine, Alternative...

Home
Family, Consumers, Cooking...

Kids and Teens
Arts, School Time, Teen Life...

News
Media, Newspapers, Weather...

Recreation
Travel, Food, Outdoors, Humor...

Reference
Maps, Education, Libraries...

Regional
US, Canada, UK, Europe...

Science
Biology, Psychology, Physics...

Shopping
Autos, Clothing, Gifts...

Society
People, Religion, Issues...

Sports
Baseball, Soccer, Basketball...

World
Deutsch, Español, Français, Italiano, Japanese, Nederlands, Polska, Dansk, Svenska...

Help build the largest human-edited directory of the web

Copyright © 1998-2006 Netscape

5,255,953 sites - 71,556 editors - over 590,000 categories

Fig. 12. ODP top level categories

A special browser is being developed which combines search engine results, the Open Directory Project (ODP) based ontology as a navigator, and search results categorization. Categories are formed based on the ODP as a predefined ontology and NWV technology is to be employed to calculate the similarity between items retrieved by the search engine and concepts in the ODP. With the interaction of users, the proposed search-browser is expected to produce more relevant search results by excluding the irrelevant, and thereby improve the quality of information returned to the user.

The ODP structure is depicted in Fig. 13.

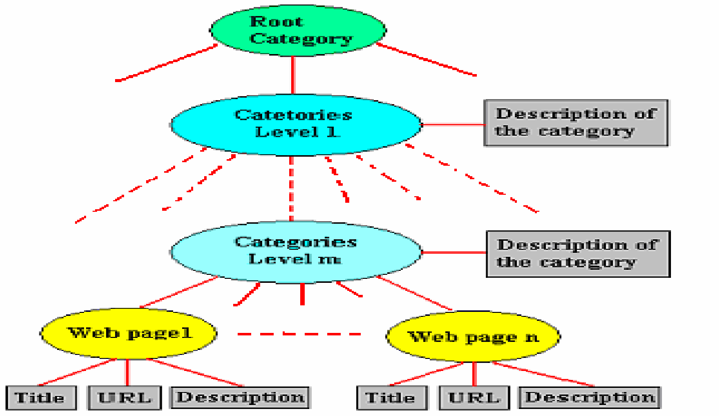


Fig. 13. Structure of the ODP

5 Summing Up

In the context of the Query-formulation → Find → Re-formulation phases of knowledge work, we set out to create some technology to provide “Assisted Query Formulation using Normalised Word Vector and Dynamic Ontological Filtering”. We have introduced and explained the Normalised Word Vector technology and via our QFR scenario exemplified its operation. The second aspect of our work, the application of dynamic ontological filtering, is in progress.

Catering for user interaction at the initial stages of the Query-formulation → Find → Re-formulation phases of knowledge work to support the creation of a Query which is truly representative of the ideas the information seeker has in mind, and to categorise the search concepts according to a suitable ontology, is expected to produce in a higher precision of search results and accommodate a diversity of information seekers, with their diversity of information needs.

References

1. Bacon, Sir Francis: Novum Organum. In Advancement of Learning - Novum Organum - New Atlantis, William Benton, Encyclopaedia Britannica, Inc. 1952. Bk.1.Sect.2. (1620)
2. Bush, V.: As We May Think. Atlantic Monthly, 176(1) July. Boston, Massachusetts (1945) 101-108
3. Dreher, H.: Interactive On-line Formative Evaluation of Student Assignments. To be presented at InSITE 2006, June 25-28, Greater Manchester, England. Online Jan (2006) http://2006.informingscience.org
4. Ide, N., Veronis, J.: Word Sense Disambiguation: The State of Art. Computational Linguistics, Vol. 24, Mar. (1998) 1-40
5. Macquarie: 2006 www.macquariedictionary.com.au, http://www.wordgenius.com.au/wordgeniusthesaurus.html
6. Open Directory Project. www.dmoz.com
7. Williams, R.: The Power of Normalised Word Vectors for Automatically Grading Essays. To be presented at InSITE 2006, June 25-28, Greater Manchester, England. Online Jan (2006) http://2006.informingscience.org
8. Williams, R., Dreher, H.: Automatically Grading Essays with Markit©. Journal of Issues in Informing Science and Information Technology Vol. 1, (2004).693-700 http://articles.iisit.org/092willi.pdf