CENTRALIZED AND DECENTRALIZED SYSTEM STRUCTURE AND EVOLUTION

By

Joao Castro, Nirav Shah, Robb Wirthlin

Professors Magee, Moses, Whitney

ESD.342 Advanced Systems Architecture

16 May 2006

INTRODUCTION

There are many kinds of systems that exhibit similar properties, but have hidden underpinnings that fundamentally differ the way they operate. The purpose of this project is to attempt to learn something fundamental about the structure of systems and the structure of the organization that controls or develops them. When a task is too complex to be carried out by a single person, a team, company or organization is put in place in order to, collaboratively, carry it out. These groups establish rules and procedures in a formal or informal way. Besides the skills of its members, the structure of an organization has a great impact on the success it has.

Our hypothesis is that systems that are structured or centrally designed are different than those that are unstructured or emerge in an evolutionary fashion.

Our approach to analyze this issue was to observe transportation networks and knowledge networks using network analysis tools and compare results to determine if any specific behavior emerged.

We focused on systems that are a result of centralized and decentralized organizations in two very different categories: knowledge and transportation. This was done to allow us to compare and control results in the comparison that are only characteristic of a single category.

Knowledge networks are established by the relationships between different topics of factual record, interest or research. Encyclopedias are an example of this, since they try to map out as much of the knowledge space as possible for reference by readers. In this work we studied in detail the Encyclopedia Britannica and the Wikipedia. We also did some research on other potential sources for this category, such as Encyclopedia Britannica online, Mathworld and Encarta [1].

The following table summarizes some of the key aspects of all these systems and organizations:

	Wikipedia	**Mathworld**	**Encyclopedia Britannica**	**Online E. B.**	**Encarta**
Established	2001	1995	1768	1994	1993
Entries	1 million articles, 340 million words	12527 articles	31,550 pages in 32 volumes 65,000 articles	120,000 articles, 55 million words; CD-ROM	41000 articles, standard 68000 articles, premium
Contributors	1 million	1	4000	4000	Not available
Stability (rate of change)	High	Med	Yearly annual update edition	Medium	Medium
Accessibility	Free	Free	Purchase	Memberships	Memberships
Peer Review	Little; becoming more frequent	Yes	Yes	Yes	Yes
Ease of change	Easy	Changes solicited; reviewed; credit given	Hard	Hard	Changes solicited; reviewed; no credit given
Depth of “knowledge”	Shallow	Mixed	Deep (Macropedia)	Mixed	Mixed

Table 1: List of various information networks [1]

Transportations networks have become integral parts of the public transit infrastructure of almost every major industrialized city in the world. Broadly speaking, the systems can be thought of as a set of interconnected networks providing different modes of transportation to the inhabitants of city. For example, in Boston, the city consists of a subway, commuter rail, busses and commuter boats and ferries. The focus of this study is the subway or 'metro' portion of larger multi-mode network. A typical subway network consists of several stations distributed throughout an urban area connected by a network of lines. To qualify as subway these lines should be located underground. Many systems however have above ground (e.g. green line in Boston) and elevated portions as well. For the purposes of this paper, systems that are largely underground are considered subways. Data was gathered in the forms maps of the subway systems. We studied six networks with the following properties:

	Year Started	Number of Stations	km of track	Planned vs. Evolved
London [2]	1863	275	415	Evolved
Beijing [3]	1965	138	197.7	Planned
Boston [4]	1897	120	101.5	Evolved
Berlin [3, 5]	1902	170	144.2	Evolved
Moscow [6]	1935	171	278.3	Planned
Tokyo [7]	1927	240	290	Planned

Table 2: Transportation networks [3, 6-9]

For both types of system we found examples of centralized and decentralized organizations running behind products in that category as illustrated by the following:

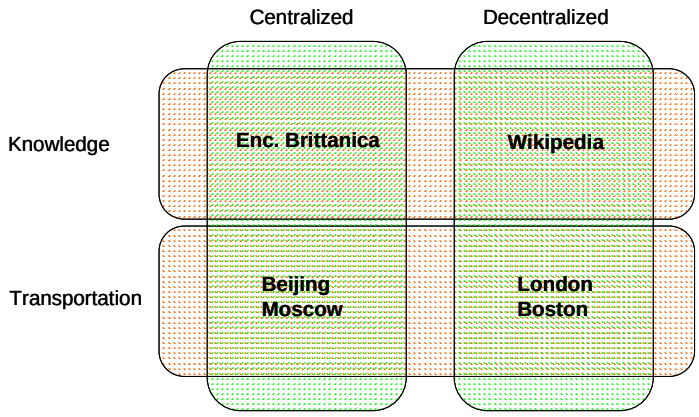


Figure 1: Definition of project space

For purposes of clarity, this paper will discuss each type of system individually, rather than in a combined fashion.

WIKIPEDIA

I. System descriptions

Wikipedia [10],[11] is an online encyclopedia that is freely accessible to consult and edit, with editions in several languages. The most popular editions are English, German and French. Contributions to the Wikipedia are generally unrestricted, with users allowed to add and change any article, as best as they

see fit. No user is ever given ownership on any of the content of the encyclopedia so future improvements are possible and most material deposited in it is considered public domain.

A. Stimulus, main actors, stakeholders / System Extent (boundary and quantities)

The Wikipedia body of knowledge is built through the effort of many different people who contribute to the several editions as volunteers.

A comparison of the most popular editions is made in the table below:

Edition	# Edits [12]	# Articles [12]	# Users [12]	# Admins [12]	Internet users [13]	% Visitors [14]	Lang. Ranking [1]
En	45587927	1026855	1091498	496	285.0M	65	4
De	14419657	358441	188709	155	53.3M	10	11
Fr	6028079	245437	77197	55	26.2M	3	18
Pl	2711067	217146	35504	48	10.6M	2	24
Ja		192000			86.3M	6	9
Sv	1750535	141060	12068	61	6.8M		68
It	2384853	140725	45873	28	28.8M	1	21
Nl	3351953	137797	28781	52	10.8M	1	40
Pt	1639353	118589	48028	31	32.0M		6
Es	2721590	96349	102130	40	45.2M	3	3

Table 3: Wikipedia language edition comparisons

The Wikipedia software and servers are maintained by the Wikimedia non-profit corporation [1]. Besides having open-content, Wikipedia runs on open-source software, effectively allowing anyone to study how it works and improve upon it for their own needs. Besides all the raw data, reasonable documentation on the structure of the system and the information is provided. The following illustration shows the technical architecture of machines used to store and display the information to users online through a mix of database, backup, web, search and proxy servers.

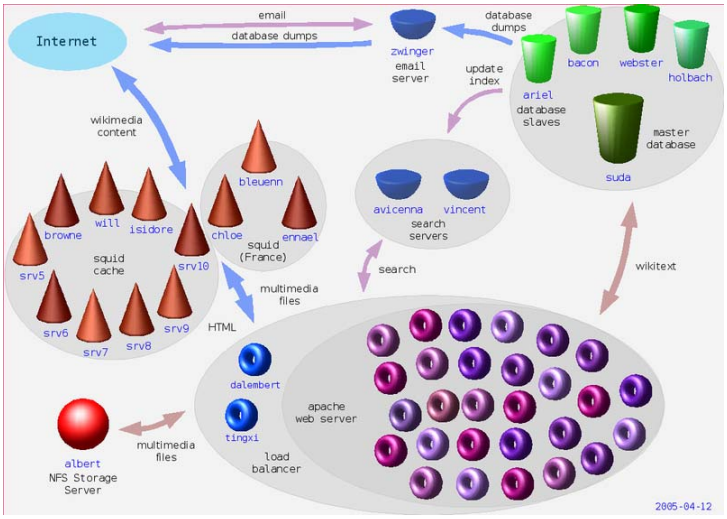


Figure 2: Wikipedia Technical Architecture [15]

B. Sources of Needs and Requirements

Wikipedia is bound only by the imaginative inputs of its unknown users and their ability to write down their thoughts. It has also been responsive to the desires of many users who have created "watch lists" to articles of interest to them, possibly to track updates, discussions, and vandalism [1].

C. System extent (Boundary and Quantities)

Wikipedia is argued by some to have no boundaries. However, some boundaries exist through the 'self-policing' that users provide, as well as the boundaries offered by the laws of science (it isn't self-aware or alive). If something is wrong in an article, someone will often correct it soon thereafter[1]. Of course, this is often a matter of opinion. Additionally, Wikipedia has 229 language editions of which 132 of them are considered "active" [1].

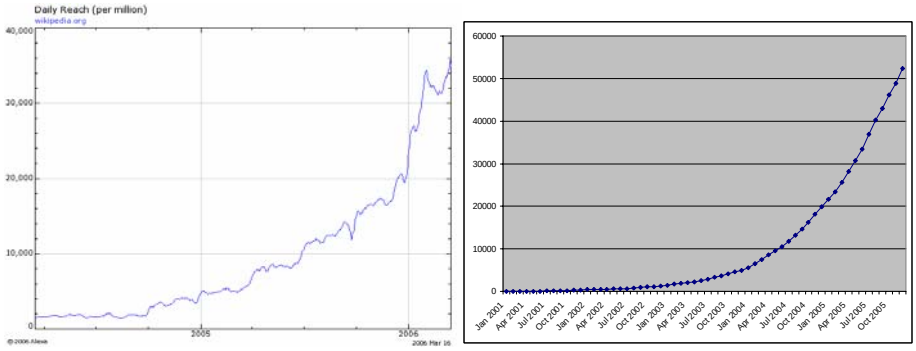
D. Mission statements, explicit if it exists or "reasonably presumed" for purposes of project

Wikipedia proclaims that it is the "free encyclopedia that anyone can edit"[1].

II. System historical background and evolution

A. History of each version fielded

The simple study of a Wikipedia dump already allows us to see a lot of the historical activity and evolution of the product. A different historical perspective on Wikipedia is to analyze its popularity and growth as a respected reference and by the number of registered users through time.



Figures 3 & 4: Traffic evolution on wikipedia.org [14] and Evolution of registered users on the Wikipedia English edition

B. Important changes in system architectural structure, defined by methods we have been discussing

Wikipedia has not had any remarkable changes in its architectural structure except in the hardware infrastructure that has hosted the material. It is built upon a scalable architecture. Please see Figure 2 above for the most recent structure layout.

C. Its size, scale, network metrics or other descriptors over time as possible

Wikipedia has a set of database files regularly copied and stored for each of the language editions. These "dumps" are publicly available at download.wikipedia.org website [12].

The English edition of Wikipedia has grown very fast, having achieved over one million articles

and the supporting files are very large. Its dump generates a compressed file over 4GB which, uncompressed, uses 452GB of disk space, making it harder to analyze without high-end resources.

Language edition	Date	Articles	Page table	Pagelinks table	Revision table	Text table
English	2006-03-03	1002326	255 MB	1.5 GB	-	-
Portuguese	2006-03-01	118589	20 MB	114 MB	-	-
Simple English	2006-03-02	7633	1.3 MB	8.5 MB	6.74 MB	224 MB

Table 4: Data about various language versions of Wikipedia [12]

The information held in Wikipedia is extensible in as much as its markup language evolves. For example, the current version does not include math operations although there are independent parties starting to explore this possibility. Once mature, this technology should be easily incorporated into Wikipedia, demonstrating its extensibility.

III. Assessment of system effectiveness over time including current critical issues

Wikipedia has been very successful. Its growth and adoption by internet users follows a pattern similar to first-to-market trends and continues with the so called "Matthew Effect" (e.g. the rich get richer). Google, whose algorithms are based upon prestige, usually offers a page from Wikipedia near the top of most search results [16].

A. Related to system characteristics like flexibility, complexity, robustness, cost, performance, etc.

As a system, Wikipedia exhibits some of the esteemed ilities:

The software tool is the key responsible for the flexibility of the content to be changed by any user, registered or not. This flexibility is also tied to resilience/repairability because any damaging change can be easily reverted.

Wikipedia demonstrates robustness, not in the software layer, but in its community of over one million users. When a user detects an error he can immediately change it. If there are issues that can't be solved immediately, users can talk to each other to solve disputes.

Finally, there are some damaging edits (erasing, defacing, and smearing) that happen in the system. Some users specialize in detecting these unwanted and unusual edits or users doing them and correct the actions of these rogue "vandals" as they are know in that community [1].

In terms of quality, since Wikipedia tries to describe the entire body of human knowledge, this aspect is hard to determine. Anyhow, Wikipedia has been favorably compared to the standard Encyclopedia Britannica and generally regarded as having similar quality [17].

The multi-language support can be regarded as a modularity attribute.

ENCYCLOPAEDIA BRITANNICA

I. System descriptions

The Encyclopaedia Britannica is widely recognized as the first modern encyclopedia and has been in publication for almost 240 years [18].

A. Stimulus, main actors, stakeholders / System Extent (boundary and quantities)

The Encyclopaedia Britannica is recognized as having a centralized approach to knowledge. The current operation of the Encyclopaedia Britannica consists of a corporate board with a board of editors. "Headquartered in Chicago, Encyclopædia Britannica, Inc. is also located in Delhi, London, Paris, Seoul, Sydney, Taipei, Tel Aviv, and Tokyo, and has already produced a variety of works in 12 languages alongside English" [19]. These editors are responsible for the overall content of the EB. In turn, there are over 4000 individual contributors to the material that makes up the EB. The contributors are usually recognized scholars in their respective fields and also considered very knowledgeable about the subject they are writing about. "The men and women of Britannica's Editorial Board of Advisors—the Nobel laureates and Pulitzer Prize winners, the leading scholars, writers, artists, public servants, and activists [who] are at the top of their fields. They meet regularly to share ideas, to debate, and to argue, in a unique collegium whose purpose is to understand today's world so that the resulting encyclopedia can be the best there is" [20]. After an article is written, the material goes through an internal peer review process, during which time the material is heavily scrutinized. Once accepted by the peer review process, the material is accepted for publication in the EB [21].

B. Sources of Needs and Requirements

Each year the EB is subject to review and possible revision. Exactly which articles are chosen for the update is not known. However, there is also a huge effort to publish the "Book of the Year" – a collection of articles and other materials that give a synopsis of the major events (newsworthy and scholarly) that occurred during the previous year [18]. Portions, or perhaps all, of this material is incorporated in the next printing of the EB. Additionally, all new material must be designated for inclusion in various areas of the EB: the Propaedia, the Micropaedia, the Macropaedia, or the Index (more discussion on these later). A reasonable supposition is that material is added, updated, or removed from the EB, depending upon its apparent impact upon the EB's Circle of Knowledge – a formal hierarchy of categorization. Additionally, in this current edition, the 15th, it has undergone major revisions at least twice: first in 1985 and then in 2005 [18].

The online version of EB, called Britannica online, was first introduced in 1994 [22]. There are several models of information access: a CD-ROM version (for purchase); a DVD version (for purchase); and the online access, with a range of information availability based upon a fee schedule [22].

C. System extent (Boundary and Quantities)

The Encyclopaedia Britannica is not much different from a boundary perspective; however, its impact is limited to English and, more recently, Spanish [22]. According to the EB Board of Editors, the EB "aspires to take all human knowledge, organize it, summarize it, and publish it in a form that people find useful…The volume of information is exploding, the world is shrinking, and digital media are changing the way we read, think, and learn" [20]. The current EB (the 15th Edition) consists of 32 Volumes. It has 31,500 pages and approximately 65,000 articles [23].

With the online portion of the EB, there are approximately 120,000 articles of more than 55 million words [23]. The CD-ROM contains more than 80,000 articles and the DVD more than 100,000 articles [23]. Many of these 'articles' are multimedia representations, whether music, pictures, movies, etc.

D. Mission statements, explicit if it exists or "reasonably presumed" for purposes of project

The mission statement for the EB is: "Who else but Britannica?" [22].

II. System historical background and evolution

A. History of each version fielded

The EB has a long history. Since its 1768 introduction, there have been 15 editions[18]. The original EB consisted of 3 volumes and now currently boasts 32 volumes. An equally interesting but perhaps more pejorative history of EB is found in Wikipedia under the search title of "Encyclopaedia Britannica" [23].

Throughout its history, there have been several prominent editors and well-renown contributors. A summarized synopsis of EB follows: The first edition had its first volume printed in 1768 [18]. The final volume (of three) was published in 1771. It was the first of its kind to consolidate important subjects into lengthy, comprehensive treatises as well as having dictionary-like entries for technical terms and other subjects [18].

The second edition was published between 1777 and 1784 [18]. The main contribution of this edition was adding some biographical articles and expanding geographic articles to include some history [18].

The 3rd through the 6th Editions are not noted for any historical contribution except that each Edition sought to correct errors found in previous editions and expand the amount of material that was covered [18]. However, a supplement to the 4th, 5th, and 6th Editions was printed between 1815 and 1824. The supplement is noteworthy because for the first time almost all of the articles were original signed contributions [18].

1830 through 1842, with the publishing of the 7th Edition, a general index was added to help find different kinds of information [18].

The 9th Edition published between 1875 and 1889 saw some turbulence as ownership of the EB passed from English to American interests [18].

The publication of the 11th Edition (1910 – 1911) saw a dramatic change in the number of articles. The traditional, lengthy, comprehensive articles were split up into more particular articles. For example, the 9th edition had 17,000 entries [18]. The 11th edition had 40,000 entries – although the total amount of text was not much different [18].

The 14th edition has several remarkable features. First, the number of articles was increased, by paring down many of the articles carried over from the 11th edition [18]. The international scope of the EB began to be seen as more than 3500 authors of all nationalities contributed to the EB [18]. Most importantly, the editorial method was changed to a continuous mode [18]. Rather than letting the EB grow

"out of date" over time and then print a new Edition, this time, the EB would be updated each year – and a "Book of the Year" would be published (and sold) to those who had previously bought a set of encyclopedias to keep their set 'up to date'.

The 15th Edition is a seminal event in the presentation, categorization, and storage of knowledge. Although published in 1974, a nearly decade long effort (and the most expensive editorial effort at that time in the world) preceded the publication in order to develop a hierarchy of knowledge called "The Circle of Knowledge" [18].

Figure 5: Simplified Circle of Knowledge [21]

Information was broken down into three sections and could also be thought of as possible layers of information. The Micropaedia (or Ready Reference), the Macropaedia (or Knowledge in Depth), and the Propaedia (the Outline of Knowledge) were born. In 1985 a major revision of the 15th Edition occurred [18]. The Macropaedia was greatly restructured and much of the information was moved into a separate two-volume Index. Furthermore, the Micropaedia and Propaedia were reorganized, revised, and redesigned.

In 1994, Britannica online was started, targeting schools and major institutions. A major revision of the entire encyclopedia database took place in 1999, the same year that Britannica.com was launched [18].

B. Important changes in system architectural structure, defined by methods we have been discussing

At the most reduceable level, the EB has had two major architectural changes since its inception. The first change EB introduced was an Index in 1815 [18]. The second significant change was The Circle of Knowledge in 1974 [18]. The idea of an Index is not new and therefore will not be discussed. However, the Circle of Knowledge is a new concept. There are three main levels in the hierarchy. First come the 10 overarching categorical "parts". Each part may have 2 to 5 "divisions" and each "division" may have 2 to 6 "sub-divisions". Finally, subject articles can tie into the Circle at any level (and these articles can also have an internal structure of several pieces). Please see the accompanying figure to see an example of this. Additionally, the printed arrangement of the EB results in a categorical layer structure consisting of summarized articles (Micropaedia) in the first 14 volumes, detailed articles (Macropaedia) in the following 15 volumes, and Propaedia (the Outline of Knowledge) in one volume (plus 2 volumes of Index materials.

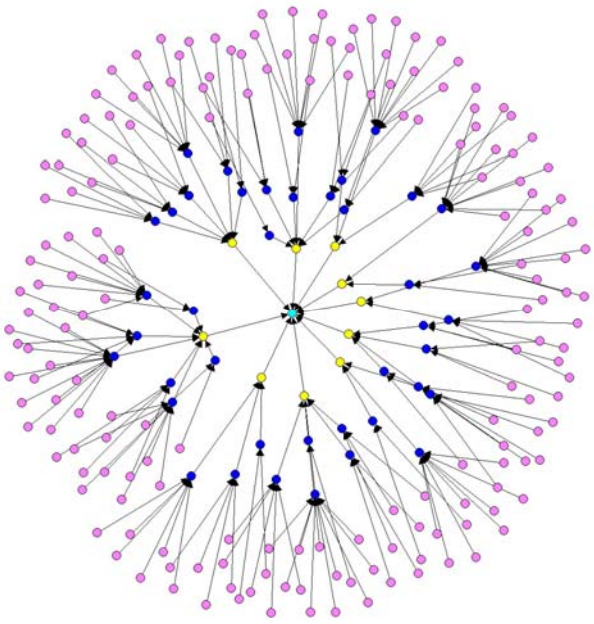


Figure 6 (left): Complete Circle of Knowledge[21]

C. Its size, scale, network metrics or other descriptors over time as possible

EB has also grown over time – as previously mentioned from 3 to 32 volumes in about 240 years and now that it is available online, the size and richness of the material online continues to grow at accelerated rates and already dwarfs the amount of the printed material. What we are not able to ascertain is if the editorial standard by which things are printed on paper is higher than what is put onto the web.

III. Assessment of system effectiveness over time including current critical issues

EB has seen an erosion of its market during the 1990s [23]. Direct sales of printed material plunged in the early 90s. It sought bankruptcy protection as it tried to cope with the accessibility and ease of the internet and the specter of instant information retrieval. Some of the actions taken by EB were to go online itself. Additionally, they purchased the "Collier's" brand as well as "Merriam-Webster" brands [23]. Now EB offers a wide range of products and services (such as Dictionaries, "Classic" literature, etc) tailored to schools, institutions, families, students, and others [22].

A. Related to system characteristics like flexibility, complexity, robustness, cost, performance, etc.

The nature of the printed Encyclopaedia Britannica offers the reader an unlimited flexibility – the

reader simply puts down a volume and picks up another or turns to another page. However, this is not entirely efficient. There is some fragility to this unless the reader is thoroughly aware of the use of the Index, the Propaedia, the Micropaedia and Macropaedia.

The online version of the EB exhibits some of the positive characteristics of Wikipedia and does not suffer from the previously listed drawbacks of the printed version.

SUBWAY NETWORKS

I. System descriptions

Each of the transportation systems studied has a unique history shaped by the time and place of its creation and development. In the interest of brevity complete histories are not provided in this paper. The reader is encouraged to consult the previously cited references. That being said, the chosen systems each have unique characteristics that make them particularly interesting for this analysis. London is the world's oldest subway system [2]. It provided an example of development that occurred without prior models. Boston plays a similar role being the oldest subway in the United States [4]. Beijing is the essential example of centrally planned subway [3, 5, 6]. In recent years, there has been a push to vastly expand the public transit capability in preparation for the 2008 summer Olympics. Given this fixed deadline, the expansion has followed far reaching yet detailed plan. This provides a good contrast to the evolved systems. Moscow is one of the world's largest systems in terms of geographic extent [3]. Berlin has the distinction of being a divided system (east and west) during the cold war period [5]. Finally, Tokyo's suburban light-rail was included for comparison to the subway systems.

A. Stimulus, main actors, stakeholders / System Extent (boundary and quantities)

A municipal transit authority such as the MBTA will govern a typical subway system. The duties and powers of the authority vary from system to system and over time. Some systems such Beijing have a powerful government authority who view the subway as part of larger set of great municipal works [6]. For others, in particular the older systems, private companies played a crucial role in the design and planning of the systems. Operations are handled by a mix of public, semi-public and private firms. The trend has been towards public/semi-public operators with distributed railways consolidating into centralized public entities. A good example of this is the emergence of the New York subway from the BMT, Independent and IRT [24]. Beyond the system managers and operators, there are other key stakeholders. These include government regulators, businesses served by the transit system, and, of course, the passengers. The relative power and influence of these varied stakeholders define the difference between the 'planned' and 'evolved' systems.

B. Sources of Needs and Requirements

Requirements for subway systems stem from the underlying transport needs of the cities they serve. In the case of subways, often inadequacy of existing infrastructure drove the creation of the subway [6]. Over time existing network of streetcars would become over-taxed or interfere too much with other

surface vehicle [6]. Subways (and L's) were a natural solution to such issues. This effect is particularly evident in the evolved systems, where integration with the existing network of streetcars, L's, and other modes of public transit often drove early development. Technological and geographic limitation also imposed requirements on subways that effected their growth. For example in Boston finding ways to cross the Charles River and go under the harbor to east Boston initially limited north and east expansion of the system [4]. Finally political realities imposed requirements that had little to do with the transport needs of the city (e.g. the division of the Berlin metro or U-Bhan) [5].

C. System extent (Boundary and Quantities)

As physical entities, subways systems have obvious boundaries denoted by the extent of their track.

D. Mission statements, explicit if it exists or "reasonably presumed" for purposes of project

Like most public works institutions, subway systems often have mission statements that speak the reliable, safe transport of their customers. For example, the MBTA mission statement reads:

"Committed to excellence, the MBTA strives to provide safe, accessible, dependable, clean, and affordable transportation to out valued customers through the dedication of our diverse and talented workforce" [25].

II. System historical background and evolution

A. History of each version fielded

Each of the transportation networks studied has a unique history. However, there do seem to be some general patterns. Many of the subways network initially operated as a part of larger network of streetcars and trams (depending on age, these cars were horse-drawn, steam or electric). This pattern is clearly seen in old cities such as London and Boston that both had established network of surface transport prior to moving underground. For example, London began with the single metropolitan line alleviating some load on a taxed surface transport system [2]. Over time, lines were added by other private companies (as many as six such lines operated at once). Eventually unification of the system occurred bringing it under a central authority.

B. Its size, scale, network metrics or other descriptors over time as possible

Though technologies come and go, the basic architecture of regularly scheduled service between stations is universal among subways and does not seem likely to change dramatically in the future. Other architectures have been proposed. Latour's book Aramis or The Love of Technology documents the unfortunate history of failed personal rapid transit system in France [26]. These PRT systems attempt to fuse the efficiency of rail with the on-demand flexibility of automobile transport. Long, mechanically coupled trains of large subway car are replaced by small (4-6 passenger) cars that can be called upon on-demand. They form virtual trains when traveling long distances and operate on a grid of tracks similar to road network.

Alas, the French attempt was plagued by significant technical and, later, political missteps leaving this potentially promising alternative architecture unrealized [26].

III. Assessment of system effectiveness over time including current critical issues

Overall, subways can be seen as successes. The metro is often one the few pieces of infrastructure in a modern city whose age surpasses 100 years. The continuing growth of cities has placed new challenges of subways. The automobile has changed the public perception with regard transport and led to a greater focus on roads over rails (e.g. the Big Dig).

A. Related to system characteristics like flexibility, complexity, robustness, cost, performance, etc.

Within the city core subway systems offer much flexibility in terms routing - there are multiple paths between destinations. As one moves to the periphery however, the system become more fragile with long branch lines that, if severed, disconnect many stations from the network. Adding connection between these branches, i.e. forming circle lines, tends to improve robustness to attack [27].

As has been stated before, the subway systems are embedded within a lager multi-modal public transit network. Latora looked at efficiency of the Boston subway both with and without the accompanying bus network. When the bus routes were included, the network was found to be more efficient [28]. The scheduled nature of subway systems tends to make them inflexible in the temporal dimension. Attempt to make so called personal rapid transit systems have met with limited (if any) success [5].

INFORMATION NETWORK ANALYSIS

Even though a complete replica of Wikipedia is available for study, current network analysis tools are not capable of dealing with networks that have over a million nodes and thirteen million edges. Also, we don't have an equivalent electronic version of the EB which we could automate tasks on. The process of studying EB is a laborious one, done manually with a 1993 edition [21].

Instead, our approach to this analysis was to select a random set of topics and then analyze the sample to compare EB and Wikipedia. The topics under analysis were selected using Wikipedia's random article function – but since Wikipedia has a much wider range of topics, we cross-checked to verify they existed in EB.

The topics were: [1, 21]

Adenomyosis; Algebra; Aluminium; Baseball; Basketball; Beekeeping; Brigadier; Cellular_automaton; Christmas; Colonization_of_Africa; Color_photography; Criminology; Design; DNA; Elisabeth_of_Bavaria; Entrepreneur; Francisco_Franco; Golf; Hans_Christian_Andersen; History_of_Manchester; Ice_cream; India; Industrial_Revolution; James_Chaney; Locomotive; Massari; Meditation; Moscow; Nobel_Peace_Prize; Paris; Politics; Population; Radio; Stradivarius; World_war_II.

We used several tools to help with the analysis of these systems: MATLAB, Microsoft Excel, UCINet and accompanying applications [29]. As previously described, EB is organized in a Circle of

Knowledge. We mapped the selected topics onto the network to produce the following:

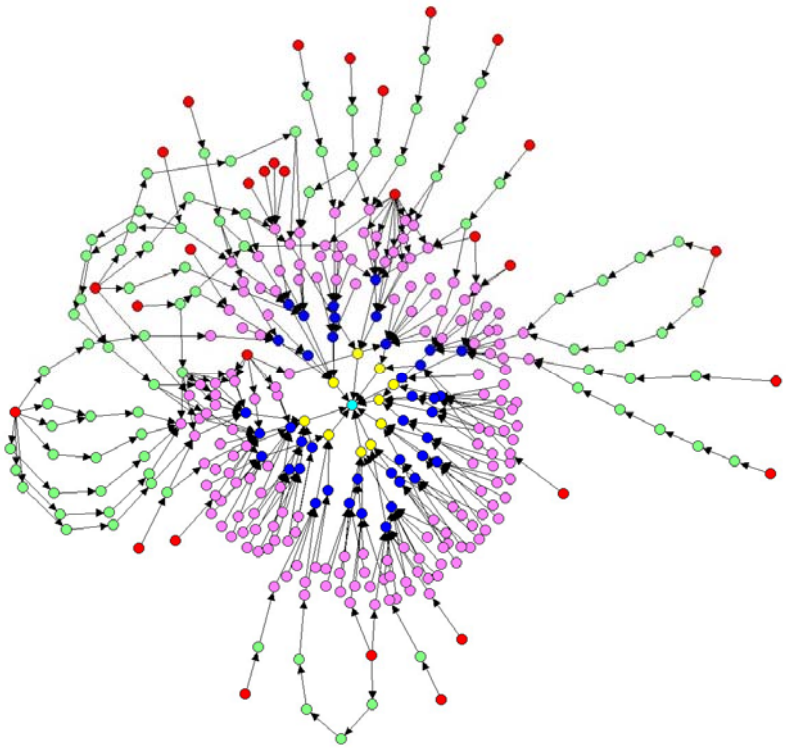


Figure 7: Circle of Knowledge with selected topics mapped to structure [21]

Red dots represent each of the topics under study. Green dots are topics that allow our sample topics to connect to the circle of knowledge, represented in purple, blue and yellow.

1. Path length in EB vs. Wikipedia

With this network, and since EB has a strict hierarchy we can now calculate the path length from each topic to another topic in our sample of EB. We also calculated the distances for the same set of topics in Wikipedia by using the online tool "Six degrees of Wikipedia" [30] and our python script (Appendix 3).

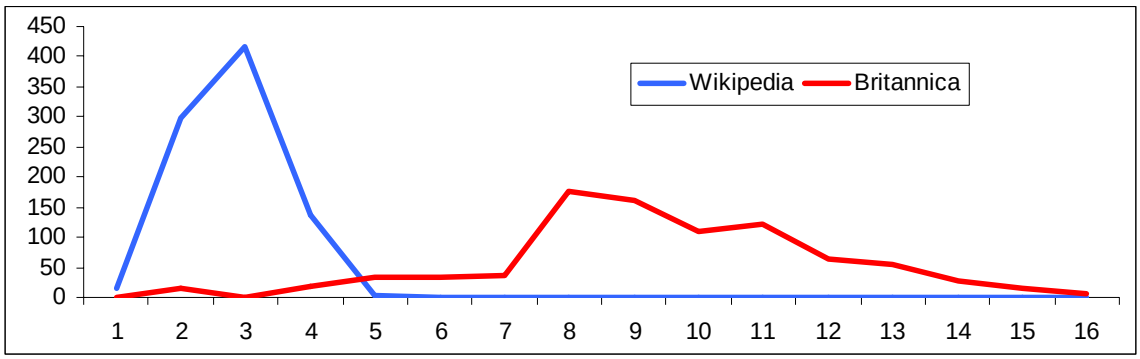


Figure 8: Histogram of Average Topic-Pair Path lengths

The average path length for EB is 9.4 and for Wikipedia is 2.8.

We can also analyze the length on a pair-by-pair basis. With this, we try to observe if the path length between two topics in one encyclopedia is similar to the path length in the other. This will allow us

to understand if a short path length in one edition also corresponds to a short distance in the other by plotting this in an EB vs. Wikipedia path length relationship plot.

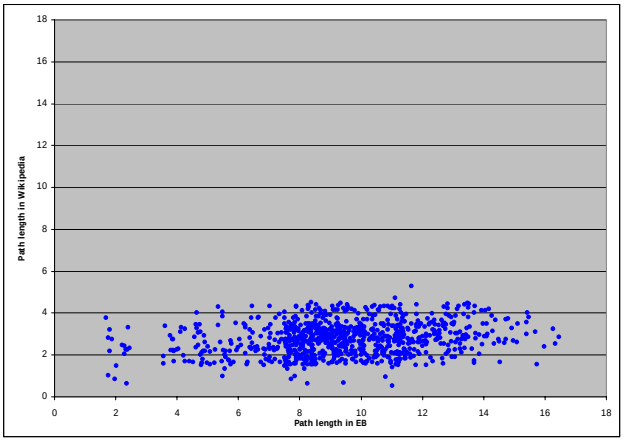


Figure 9: Wikipedia Topic-Pair Path Lengths vs. EB

From this graph we can observe that Wikipedia has a consistent path length from topic to topic, even for the pairs that rank as having a high path length in EB. Distance between topics is given by an integer value but in order to highlight the number of pairs studied, we introduced a small random displacement (<1) in the values to produce the above cloud of dots.

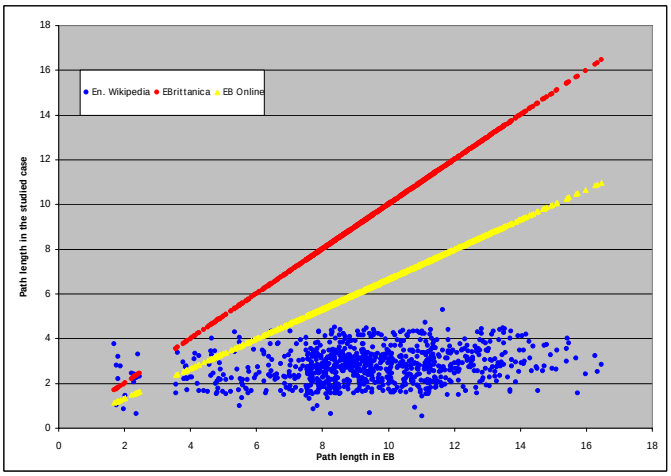


Figure 10: Addition of EB Printed Topic-Pair Path Length & Predicted Online EB Topic-Pair Path Length

We then also plotted EB on the same graph (an obvious slope 1 line) to further illustrate the difference between EB and Wikipedia. We also speculate that if were able to repeat the exercise for the online edition of EB we would obtain a plot similar to the yellow line above. EB online will have a shorter path length than EB because it can include direct links between articles, not having to use the hierarchy of the circle of knowledge to connect two topics. The yellow line would also have a higher path length than Wikipedia since several paths in Wikipedia use pages related with dates, something not expected in a pure knowledge relationship.

2. Impact of hierarchy

Using the same path length data between topics we can analyze if there are topics that have a tendency to be closer to others, adequately mapping out the knowledge space. In order to analyze a multi-dimensional space we used dendrograms and obtained the following results:

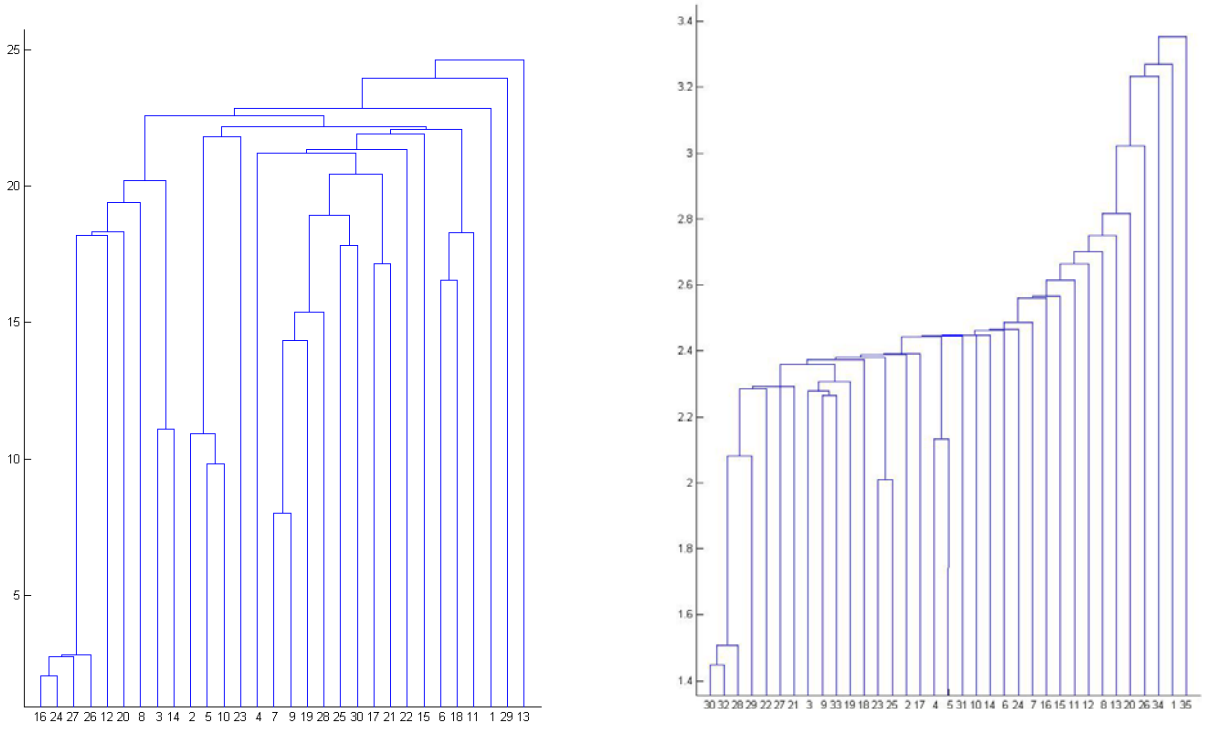


Figure 11: Dendrograms of EB (left) and Wikipedia (right)

These two graphs represent the distances between topics in a 24 and a 23 dimensional space for the EB (on left) and Wikipedia (on right) representations, respectively.

From them we can see that EB's hierarchy forces a clustering of topics that are positioned in the same branches while Wikipedia blurs this distinction. The differences in the dendrograms are a clear illustration of the random nature of connections in Wikipedia, with very small distances between topics, comparatively to the whole group of topics.

Also of note is the very close cluster between the topics "History of Manchester", "Paris", "Moscow" and "Population" in both cases. This is a clear indication that, despite not being able to make distinctions between very different areas, Wikipedia has the capability of clustering very similar topics together.

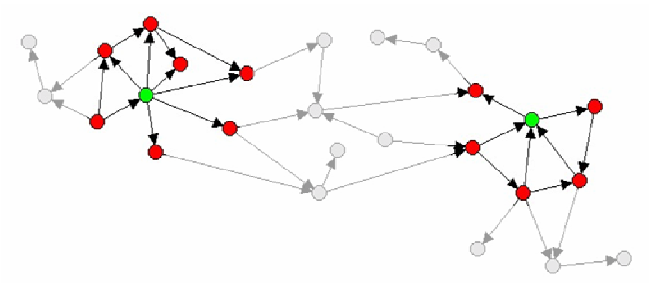
3. Historygrams

Neither EB nor Wikipedia are frozen, immutable study objects. EB has been changing for over 250 years and Wikipedia, despite much younger, has also been changing very fast, with the English edition having over 45 million registered edits in five years.

A better understanding of how these two systems change through time is essential to also understand the differences that a centralized and a decentralized organization structure can impact the

output. We already discussed some historical aspects of both systems and now we present a quantitative method to analyze the local impact of a topic.

First, we discover the topic's horizon network which consists of all the nodes that the topic links to and the topics that link to the topic, as illustrated here:



We then determine the date when each of the articles in the horizon was started and plot a histogram of those dates. This method is easy to use in Wikipedia since it stores this type of details but is not in EB.

Figure 12: Article Horizon depiction

Because we mark the topic under study differently in the histogram we call these plots "historygrams".

The following figure shows 30 historygrams of articles in the simple English Wikipedia. The red triangle marks the time when the central topic was created.

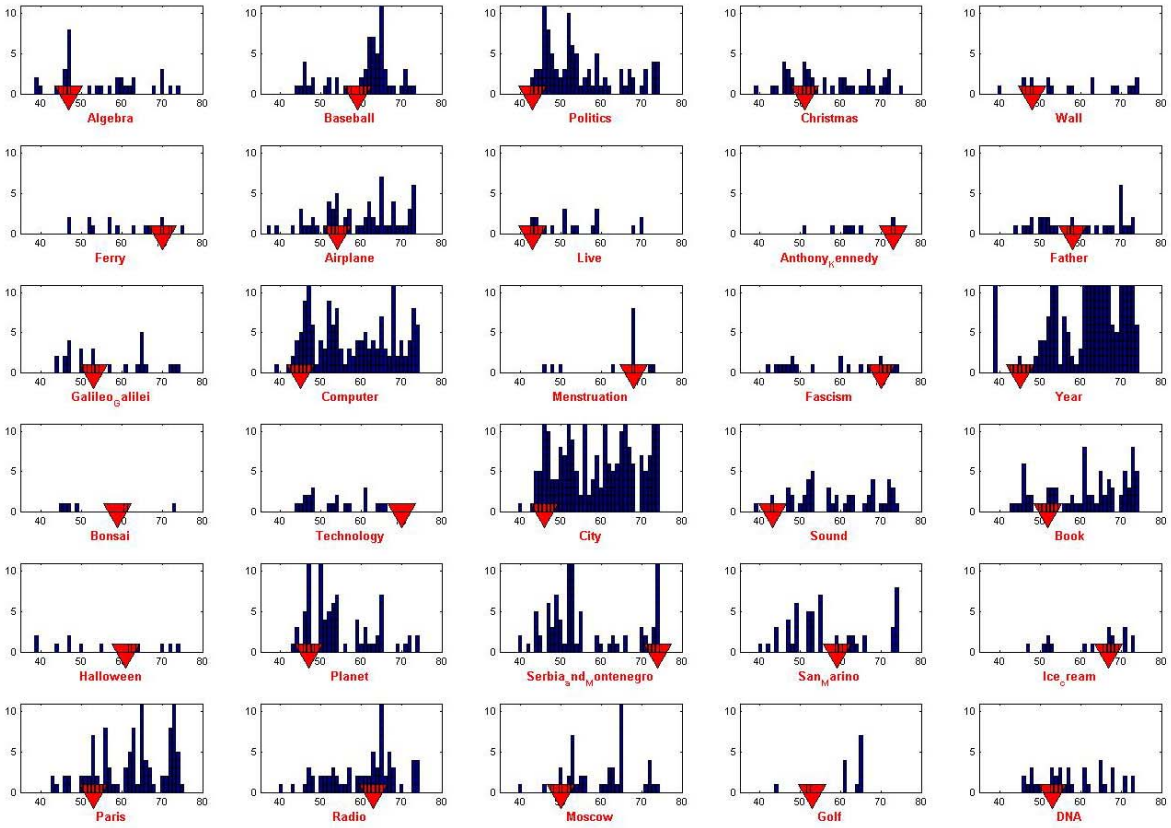


Figure 13: Historygrams of Randomly Selected Topics in Wikipedia

As we can see and calculate from these plots, 80% of the horizon articles are created after the original topic in the horizon. This shows that each new topic is responsible for creating several more in the knowledge space, effectively promoting the growth and expansion of Wikipedia which can also be seen in

the article count graph in figure 14.

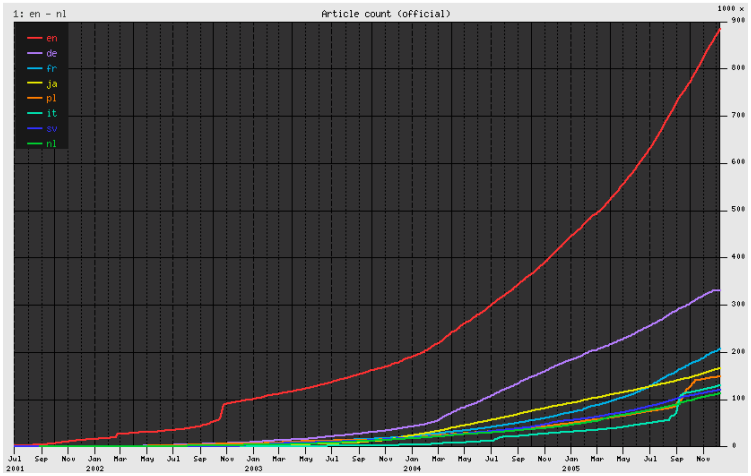


Figure 14: Wikipedia Total Article Count by top languages [31]

4. Timeline of key events

Another way of observing the evolution of both knowledge systems is by understanding and comparing some of the key events EB and Wikipedia have gone through. An interesting aspect of this analysis is to verify that both systems have gone through similar events, almost as if they were milestones in the development of a knowledge repository or encyclopedia.

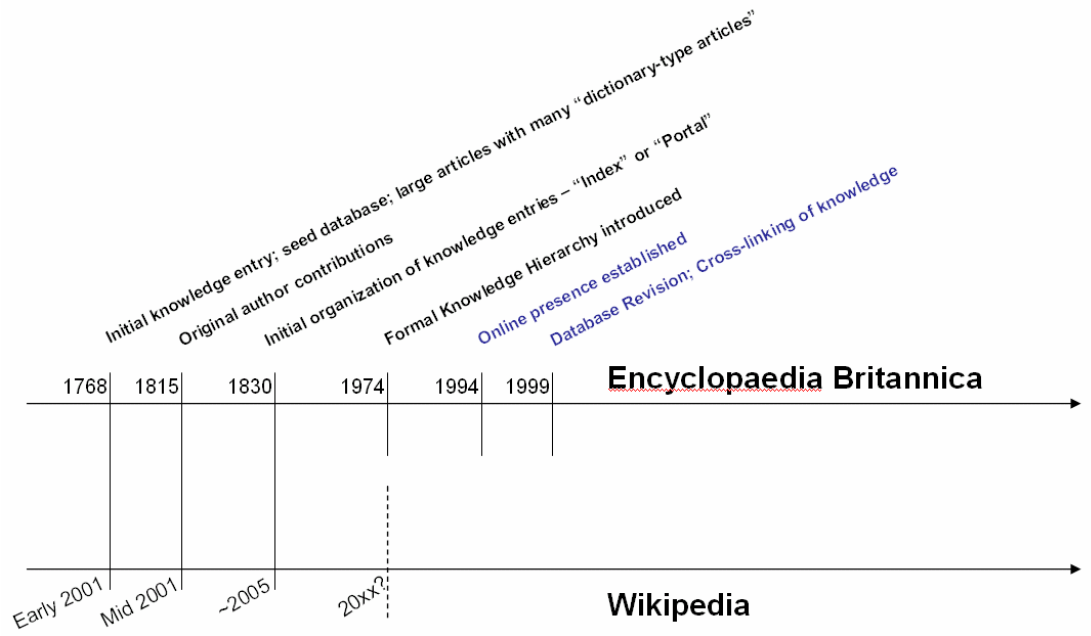


Figure 15: Timeline comparison

One other interesting aspect of this graph is that it clearly illustrates the different speeds in evolution of the two systems. From our understanding, the different clockspeeds are attributable to the development of technology, and specifically to the internet and web platform on which Wikipedia operates.

SUBWAY SYSTEM ANALYSIS

Network metrics were computed for the subway and light-rail systems [29]. No noticeable differences were found between the two groups:

	n	m	<k>	C1	C2	L	R
London	92	139	3.02	0.222	0.1595	5.394	0.0997
Beijing	29	82	2.83	0.237	0.0667	3.409	-0.1053
Boston	21	44	2.09	0.074	0.0317	3.562	-.3011
Berlin	75	222	2.96	0.117	0.0812	5.164	0.0957
Moscow	51	164	3.21	0.106	0.0791	3.916	0.1846
Moscow (w/ rail)	136	408	3.00	0.080	0.0591	6.037	0.2601
Tokyo	147	204	2.77	0.078	0.0522	6.432	-0.0911

Table 5: Network Metrics for Selected Subway Systems

Centrality metrics also revealed no interesting differences between planned and evolved systems.

COMBINED ANALYSIS

Potential architectural improvements: Visualizing change in the systems using the DWS chart.

Duncan, Watts and Sabel proposed a way to categorize hierarchical systems according to two parameters, depth and breadth of links across branches[32]. We adapted their model to the systems we studied and plotted their movements on the chart through time to illustrate differences.

EB started as a collection of articles. In 1842, with the introduction of the index, all articles had a link to that single element. This represents a hierarchy with minimal depth and no breadth of links and so we can make its first representation on the chart. In 1974, with the introduction of the circle of knowledge, EB gained a deeper hierarchy while maintaining a very low cross-linkage between articles of different branches. Finally, EB started to take advantage of the online technology and introduced links between articles of different branches.

Wikipedia also started as collection of articles with no hierarchy, but with a very high degree of linkage between them. In 2005, the community that edits Wikipedia started to categorize entries into what is starting to shape up as a hierarchy[1]. It is still an on-going process so we tentatively draw wikipedia's evolution on the chart as moving to a more structured system.

The subway system can also be interpreted using the DWS framework. Most systems after some years of existence develop into a core-periphery structure with a central high interconnected core with branches going out to the suburbs. See, for example, London in the 1920s.

Image removed for copyright reasons.
1921 map of London subway.
http://www.clarksbury.com/cdl/maps/tube21.jpg

Figure 16: London Subway System Map, Early 1920s [8]

Once this basic star pattern in established two forces govern the evolution of the system. One is the desire for citizens to travel between more developed suburbs bypassing the core. This leads to circle lines that appear as interconnection between hierarchies (RID on the DWS chart). The other forces, is local interconnection and expansion of the branches into networks in there own right. This leads to a layered structure. When combined, these forces tend to make the system multi-scale with both layers as well as rings. While the maps of London underground [8] do seem to show this pattern of growth formal test of the hypothesis require network representation of the full history of the underground. Given that the maps are available, future student groups should attempt to digitize them and numerically test growth hypotheses such as the one proposed above.

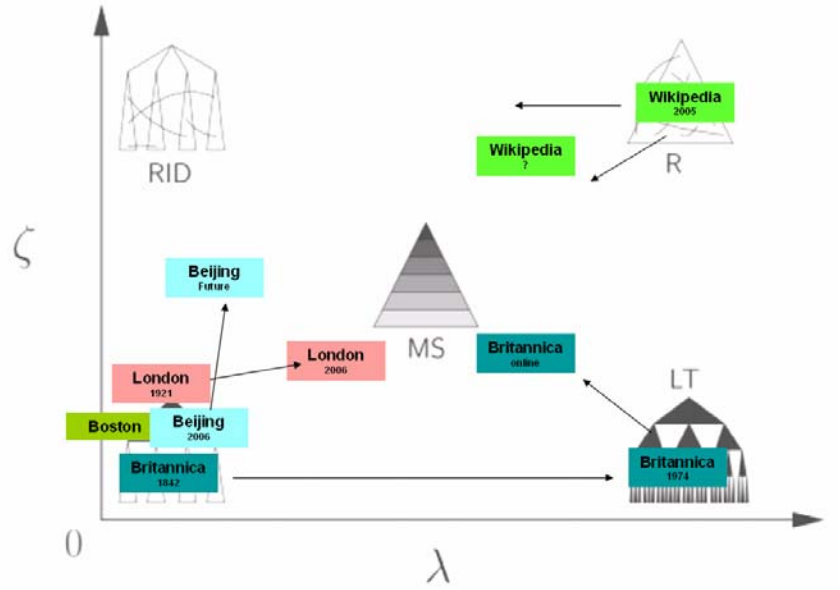


Figure 17: Using DWS framework as unifying construct for Studied Systems

CONCLUSIONS

Reflections and comparisons (what worked and didn't)

This project was a huge undertaking. Upon close inspection, the motivation and idea for this project is essentially to compare and contrast at least three systems using network tools, ideas, and methodologies. Based upon the supposition that the genesis of these systems, centralized or decentralized (planned or evolved), would lend themselves well to network theory analysis, we went to work. However, the nature of some of these systems precluded the use of the tools that we would have liked to use.

For instance, the printed Encyclopaedia Britannica doesn't really have network statistics readily available; except for those we inferred or derived by hand. The online version of the EB doesn't make available the network properties of the system, so we were limited to our own limited observations. Wikipedia was truly the only information system that could provide all of the data to use for a full network analysis, but because of its size, we were limited to the machine and computational capability of the network analysis tools we had.

Analyzing the subway networks from a planned vs. evolving state did not yield the kinds of information we were hoping to see, especially the ideas of betweeness, centrality, and path length did not necessarily pan out.

Therefore, we had to develop a different and original approach to understand these systems. We feel that we were successful in doing this. For the case of the knowledge networks, we were able to look at paired-topic path length comparisons to see if the structures behind these knowledge systems imposed a "penalty" on path length: it does. However, we also saw that computational reality and end user "search habits" were likely unaffected by any kind of imposed hierarchy.

Further, transportation systems are subject to long-term constraints such as legacy and geography that make changes slow. Therefore, riders on the subway system are unlikely to perceive any kind of network penalty based upon the layout of the system.

Analogies to other systems or kinds of systems

Although the DWS framework presupposes an organizational hierarchy, we found that it stimulated our thinking and allowed us to draw analogies to it. For instance, the paper purports that the multi-scale structure is the most preferred of forms. In observing the evolution of these systems, we see a migration or convergence to an analogous common structure. We also acknowledge there are various constraints that prevent these systems from ever truly reaching an idyllic form, but the idea that hierarchies emerge for knowledge structures and that in transportation networks, circle structures and trees become ever more interconnected, is very compelling. We hope there will be further studies to pursue these questions.

Similarly, while the DWS presupposes an existing hierarchy in its examination, our systems did not start with one. In the world of social networks, new businesses and ventures may not start with any formal structure, but evidence suggests that hierarchies emerge over time. Only after these structures emerge is the DWS framework appropriate to be used. Therefore, there is a gap in the available knowledge and tools available to study systems, whether informational networks, transportation systems, or even organizations, in their most nascent stages. This would be another fruitful area of research.

What did you learn by doing this project

Network metrics are not necessarily the panacea for all systems nor do all systems that exhibit network structures and behaviors work well using existing network analysis tools. However, the ideas behind network theory are compelling and suggest that this field is only emerging. As time goes on, further extensions and additions to theory, as well as the development of more robust tools will enrich the studies

of networks of all kinds.

Network analysis is also hampered by its static nature. Our systems, particularly the informational network systems, are dynamic and the conclusions we have drawn in this paper are likely already outdated.

What aspects of the topic area did you find strong or weak, more or less useful, etc.

Classical network analysis and theory relies upon abstractions and we feel that this actually causes you to lose valuable and useful information. Our observation is that this means this way of representation has been taken to its furthest limit and new representations are needed. We need to develop methods for "networks of networks", for instance, we can analyze a subway network, or a bus network. How do we analyze both overlaid upon the street network? Furthermore, there is the network of authorship (or contributions) in the encyclopedias overlaid on the articles within those encyclopedias. How can these be analyzed simultaneously? Our tools and theories are very limited or can not yet accommodate these questions.

Growth models seem to give insights into fundamental mechanisms underlying network dynamics. (If you can recreate the network through your model, you should be able to understand how it works – in other words, it is testable). We know such a paper exists for commuter rail systems [33], but does not for information networks.

Summary

Information networks and subway systems seem to be converging to common analogous structures despite different underlying structure and philosophies. Network theory and tools are limited in application to systems such as these for reasons listed above and shows an area for great research opportunities.

APPENDIX A: BIBLIOGRAPHY AND NOTES

1. unknown. *Wikipedia*. 2006 [cited 2005 April]; Available from: www.wikipedia.org.

2. Howson, H.F., *London's Underground*. 2nd ed. 1986, London: Ian Allan Ltd. 160.

3. unknown. *UrbanRail*. 2006 [cited 2006; Available from: www.urbanrail.net.

4. Cudahy, B.J., *Change at Park Street Under: The Story of Boston's Subways*. 1972, Brattleboro, VT: The Stephen Greene Press. 64.

5. *Wikipedia*.

6. Bodrick and Benson, *Labyrinths of Iron: A History of the World's Subways*. 1981: Newsweek Books.

7. Kato, S., *Development of Large Cities and Progress in Railway Transportation.* Japanese Railway & Transport Review, 1996(8): p. 44-48.

8. unknown. *London Tube Maps*. 2006 [cited 2006; Available from: www.clarksbury.com/cdl/maps.html.

9. Whitney, D., *Subway Archive*. unk, MIT.

10. Cunningham, W., *What is wiki?* 2004.

11. A Wiki is a piece of server software that allows users to freely create and edit Web page content using any Web browser. It allows users to edit the content on a webpage via a form, with a simplified notation and markup language. This way the user does not need to have advanced technical skills to contribute.

12. unknown. *Wikipedia statistics*. 2006 [cited 2006 April]; Available from: www.wikipedia.org, http://download.wikipedia.org.

13. unknown. *Internet Statistics*. 2006 [cited 2006 April]; Available from: www.internetworldstats.com.

14. unknown. *Visitors*. 2006 [cited 2006 April]; Available from: www.alexa.com.

15. unknown. *Technical Architecture of Wikipedia*. 2006 [cited 2006 April]; Available from: www.mediawiki.org.

16. unknown. *Wikipedia*. 2006 [cited 2006 15 May 2006]; Wikipedia Search]. Available from: www.google.com.

17. Giles, J., *Internet encyclopedias go head to head*, in *NATURE*. 2005. p. 990-901.

18. Encyclopaedia Britannica. *Encyclopaedia Britannica*. 2006 [cited 2006 May 9]; Available from: http://www.search.eb.com/eb/article-2111.

19. unknown. *International Publishing*. 2006 [cited 2006 May]; Available from: http://corporate.britannica.com/intl_publishing.html.

20. Encyclopaedia Britannica. *Editorial Board*. 2006 [cited 2006 May]; Available from: http://corporate.britannica.com/board/.

21. Encyclopaedia Britannica, *Encyclopaedia Britannica*. 15th ed. Vol. all. 1993, Chicago: Encyclopaedia Britannica.

22. Encyclopaedia Britannica. *Encyclopaedia Britannica Online*. [Internet Web pages] 2006 [cited 2006; Available from: www.britannica.com.

23. unknown. *Encyclopaedia Britannica*. 2006 [cited 2006 April]; Available from: www.wikipedia.org.

24. unknown. *NYC Subway*. 2006 [cited 2006 April]; Available from: www.nycsubway.org.

25. MBTA. *The MBTA Organization*. 2006 [cited 2006 May]; Available from: www.mbta.com/insidethet/budget.asp.

26. Latour, B., *Aramis or The Love of Technology*. 1996, Cambridge: Harvard University Press. 336.

27. Latora, V. and M. Marchiori, *Vulnerability and Protection of Critical Infrastructures*. 2006, Universita' di Catania

Massachusetts Institute of Technology. p. 4.

28. Latora, V. and M. Marchiori, *Efficient Behaviour of Small-World Networks.* Physical Review Letters, 2001. **87**(19): p. 4.

29. Borgatti, S.P., M.G. Everett, and L.C. Freeman, *Ucinet for Windows: Software for Social Network Analysis*. 2002, Analytic Technologies: Harvard, MA.

30. unknown. *Six degrees of wikipedia*. 2006 [cited 2006; Available from: http://tools.wikimedia.de/sixdeg.

31. unknown. *Article count (official)*. 2006 [cited 2006 April]; Available from: http://stats.wikimedia.org/EN/PlotsPngArticlesTotal.htm.

32. Dodds, P.S., D.J. Watts, and C.F. Sabel, *Information exchange and the robustness of*

organizational networks. PNAS, 2003. **100**(21): p. 12516-12521.

33. Gastner, M.T. and M.E. Newman, *Shape and efficiency in spatial distribution networks.* Journal of Statistical Mechanics: Theory and Experiment, 2006(1): p. 4.

APPENDIX B - WIKIPEDIA TESTBED

Wikipedia has a software architecture layered with the following appearance:

User Layer	Web clients		
Network Layer	Apache Webserver		
Logic Layer	MediaWiki's PHP scripts		
	PHP		
Data Layer	File System	MySQL Database	Caching System

All the content is stored in a database, so, in order to analyze this data we need to know the information structure used in that database. Of all the files, the best suited to study how the information is organized is the Page, Pagelinks and Revision tables [2].

The relevant structure of these tables is as follows:

Pagelinks – link relationship between pages

Field	Description
pl_from	ID number of the page that has a link
pl_namespace	Type of page linking
pl_title	Title of the destination link
…	

Page – information for a specific page

Field	Description
page_id	Uniquely identifying key
page_namespace	Type of page linking
page_title	Title of the page
page_counter	Number of times this page has been viewed
page_touched	timestamp whenever the page changes
page_latest	key to revision.rev_id of the current revision
page_len	Uncompressed length in bytes of the page's current source text
…	

Revision - information on the authors of each edit

Field	Description
rev_id	Uniquely identifying key
rev_page	Page to which this revision pertains
rev_text_id	Key to text.old_id in the "Text" table, where the actual bulk text is stored
rev_user	Key to user_id of the user who made this edit
rev_user_text	Text username or IP address of the editor
rev_timestamp	Time of edit
…	

To install a local and simplified version for study, some specific software tools are required, namely:

- BZ2 unpacker
- XML to SQL file converter, we used xml2sql-0.5-win32)

- Local installation of the MySQL Server 5.0 and the MySQL Query Browser

- MySQL Database Connector for Matlab, we used Jörg Buchholz's "Myblob" connector

- Matlab to run the scripts

Wikipedia has a set of database files regularly copied and stored for each of the language editions. These "dumps" are publicly available at download.wikipedia.org website. We used the following dumps to conduct our study:

English

Simple English

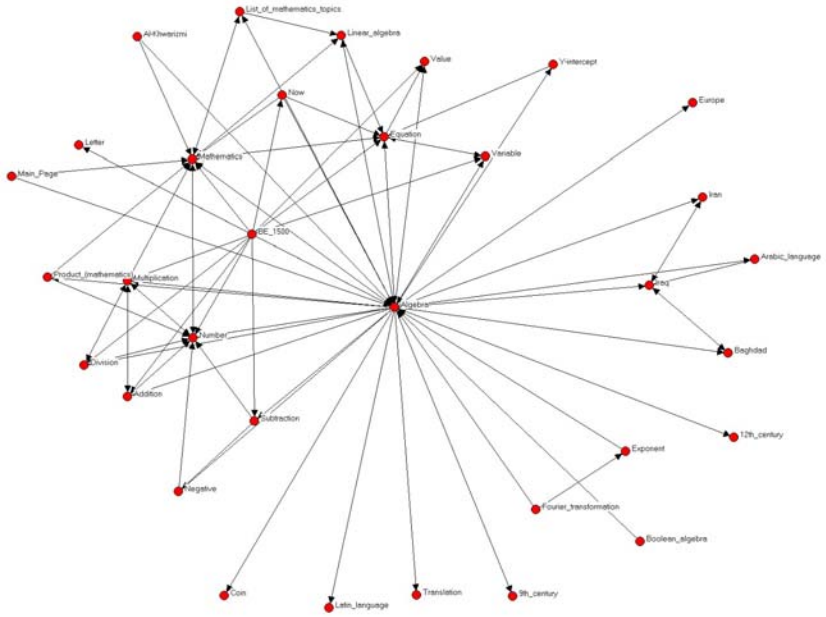
APPENDIX C - MATLAB SCRIPTS

These scripts interact with the Wikipedia testbed by using Jörg Buchholz's "Myblob" connector. This allows Matlab to talk with the MySQL database in a very simple way.

Horizon builder

This script queries the database for all articles linked by the topic and all articles that link to the topic. After obtaining this list the script then goes through each of the articles in that list to query for links to any other of the articles in the full list.

This information is written to a .VNA file in the Matlab work area. Example of horizon with topic 'Algebra' on the Simple English edition:



Source code:

```
% Joao Castro - joaoc@mit.edu - copyright 2006
%
% Gets the Data from a Mysql database with WikiMedia 1.5 data structure
% and builds the inbound and outbound link diagram in .VNA file format
%
% This DOES NOT WORK WELL with databases that have non-ASCII characters
%
% USAGE: Alter the line below with the inputs to the topic you want to
% analyze and then run the M-file
%
% REQUIREMENTS: the Wikipedia installation must have the following tables:
%   - page
%   - pagelinks
%
%=========================================================
% INPUTS
```

```
% Insert your own mySQL database parameters here:
user = 'username';
password = 'password';
server = 'localhost'; %localhost for if you're running the database on the
% same machine as Matlab or IP address to reach remote MySQL server
database = 'Wikipedia';

% Topic to use as center-node
topic = 'Algebra';

%============================================================
% Create file with topic
fid = fopen([topic,'.vna'], 'w');
fprintf(fid, '*tie data\nfrom\tto\tlinks\n');

% inbound links
connection = myblob_open (user, password, server, database);
command = ['SELECT page.page_id, page.page_title FROM page, pagelinks WHERE
page.page_id = pagelinks.pl_from AND pagelinks.pl_title="', topic,'" AND
page.page_namespace=0;';];
inbound_array = myblob_command (connection, command);
myblob_close (connection);

%go through the collected data and pass it to the file
a=size(inbound_array);
b=a(1,2);
for a = 1:b
    stringtofile = [cell2mat(inbound_array(2,a)),'\t',topic,'\t1\n'];
    fprintf(fid, stringtofile);
end

%============================================================
%outbound links
connection = myblob_open (user, password, server, database);
command = ['SELECT page.page_id, pagelinks.pl_title FROM pagelinks, page WHERE
pagelinks.pl_from= (SELECT page_id FROM page WHERE page_title ="',topic,'" AND
page_namespace = 0)AND page.page_title = pagelinks.pl_title AND page.page_namespace
= 0 ORDER by pagelinks.pl_title;'];
outbound_array = myblob_command (connection, command);
myblob_close (connection);

%go through the collected data and pass it to the file
a=size(outbound_array);
b=a(1,2);
for a = 1:b
    stringtofile = [topic,'\t',cell2mat(outbound_array(2,a)),'\t1\n'];
    fprintf(fid, stringtofile);
end

%============================================================
%check relationships between articles in topic horizon
%we first build the horizon_array and then check for outbound
%links in that horizon that connects to any other page in the
%horizon group.

%build the horizon query
%NEW horizon_array builder. no repeated entries!
```

```
horizon_array = outbound_array;
a=size(inbound_array);
b=size(outbound_array);
for i=1:a(1,2)
    repeated =0;
    for j=1:b(1,2)
        if cell2mat(inbound_array(1,i)) == cell2mat(outbound_array(1,j))
            repeated =1;
        end
    end
    if repeated == 0
        horizon_array = [horizon_array,inbound_array(:,i)];
    end
end

a=size(horizon_array);

%generate the OR part of query with list of the pages that are in 'Topic' horizon
clear substring;
substring = ['(page.page_id=',num2str(cell2mat(horizon_array(1,1)))];
for subindex=2:a(1,2)
        substring = [substring,' OR
page.page_id=',num2str(cell2mat(horizon_array(1,subindex)))];
end
substring = [substring,') ORDER by page.page_id;'];

%generate the query for each one of the pages on the horizon, to check for
%links to other horizon pages.
for seed=1:a(1,2)
    command = ['SELECT page.page_id, pagelinks.pl_title FROM pagelinks, page WHERE
pagelinks.pl_from=', num2str(cell2mat(horizon_array(1,seed))),' AND page.page_title
= pagelinks.pl_title AND page.page_namespace = 0 AND ',substring];

    connection = myblob_open (user, password, server, database);
    horizon_list = myblob_command (connection, command);
    myblob_close (connection);

    a=size(horizon_list);
    if a(1,2)>0
        for i=1:a(1,2)
            stringtofile =
sprintf('%s\t%s\t1\n',cell2mat(horizon_array(2,seed)),cell2mat(horizon_list(2,i)));
            fprintf(fid, stringtofile)
            clear stringtofile;
        end;
    end;
end;

%============================================================
fclose(fid);
```

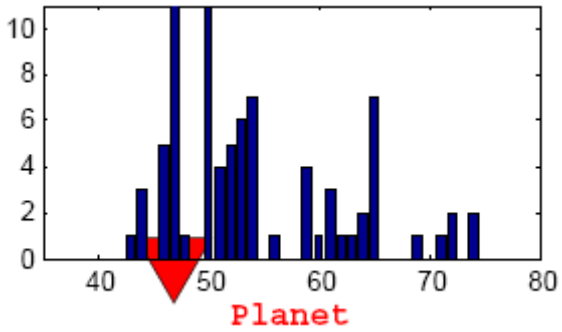
Historygram builder

The historygram script queries the database for all articles linked by the topic and all articles that

link to the topic. After obtaining this list the script then goes through each of the articles and obtains the date that the article was first created.

This information is written to a .CSV file file in the Matlab work area and is plotted in a historygram inside matlab. For a better explanation of what a historygram is, refer to the main body of this document.

Example of a historygram for the article 'Planet' on the Simple English edition:



Source code:

```
function history(topic,sub)

% Joao Castro - joaoc@mit.edu - copyright 2006
%
% Gets the Data from a Mysql database with WikiMedia 1.5 data structure
% and builds the inbound and outbound link list and
% 1) outputs the number of hours since 2000 in a CSV file
% 2) plots the historygram
%
% USAGE: topics must be passed as strings in first parameter.
% Second parameter is position of historygram in a multiplot
% EXAMPLE:
%          history('Algebra',1)
%          history('Baseball',2)
%
% REQUIREMENTS: the Wikipedia installation must have the following tables:
%   - page
%   - pagelinks
%   - revision

%============================================================
% INPUTS
% Insert your own mySQL database parameters here:
user = 'username';
password = 'password';
server = 'localhost'; %localhost for if you're running the database on the same
machine as Matlab or IP address to reach remote MySQL server
database = 'Wikipedia';

%============================================================
% inbound links
connection = myblob_open (user, password, server, database);
command = ['SELECT page.page_id, page.page_title FROM page, pagelinks WHERE
page.page_id = pagelinks.pl_from AND pagelinks.pl_title="', topic,'" AND
```

```
page.page_namespace=0;';];
inbound_array = myblob_command (connection, command);
myblob_close (connection);

%go through the collected data and pass it to the file
%a=size(inbound_array);
%b=a(1,2);
%for a = 1:b
%    stringtofile = [cell2mat(inbound_array(2,a)),'\t',topic,'\t1\n'];
%    fprintf(fid, stringtofile);
%end

%============================================================
%outbound links
connection = myblob_open (user, password, server, database);
command = ['SELECT page.page_id, pagelinks.pl_title FROM pagelinks, page WHERE
pagelinks.pl_from= (SELECT page_id FROM page WHERE page_title ="',topic,'" AND
page_namespace = 0)AND page.page_title = pagelinks.pl_title AND
page.page_namespace = 0 ORDER by pagelinks.pl_title;'];
outbound_array = myblob_command (connection, command);
myblob_close (connection);

%go through the collected data and pass it to the file
%a=size(outbound_array);
%b=a(1,2);
%for a = 1:b
%    stringtofile = [topic,'\t',cell2mat(outbound_array(2,a)),'\t1\n'];
%    fprintf(fid, stringtofile);
%end

%============================================================
%check relationships between articles in topic horizon
%we first build the horizon_array and then check for outbound
%links in that horizon that connects to any other page in the
%horizon group.

horizon_array = outbound_array;
a=size(inbound_array);
b=size(outbound_array);
for i=1:a(1,2)
    repeated =0;
    for j=1:b(1,2)
        if cell2mat(inbound_array(1,i)) == cell2mat(outbound_array(1,j))
            repeated =1;
        end
    end
    if repeated == 0
        horizon_array = [horizon_array,inbound_array(:,i)];
    end
end

%get the timestamp of the topic
connection = myblob_open (user, password, server, database);
command = ['Select rev_timestamp from revision where rev_page=(SELECT page_id
FROM page WHERE page_title ="',topic,'" AND page_namespace = 0) order by
rev_timestamp asc limit 1;'];
topic_first_date = myblob_command (connection, command);
```

```
myblob_close (connection);

%get timestamps of horizon
a=size(horizon_array);
loop = a(1,2);

for i=1:loop
    connection = myblob_open (user, password, server, database);
    command = ['Select rev_timestamp from revision where rev_page=',
num2str(cell2mat(horizon_array(1,i))),' order by rev_timestamp asc limit 1;'];
    first_date = myblob_command (connection, command);
    myblob_close (connection);

    horizon_array(3,i) = {cell2mat(first_date)};
end

%Aprox number of hours since 2000
date=cell2mat(topic_first_date);
year=str2num(sscanf(date,'%c',4));
month=str2num(sscanf(date,'%c',6))-year*100;
day=str2num(sscanf(date,'%c',8))-year*10000-month*100;
hour=str2num(sscanf(date,'%c',10))-year*1000000-month*10000-day*100;

topicHours2000=(((year-2000)*12+month)*30+day)*24+hour;
topicMonthSince2000=(((year-2000)*12+month));

%make CSV file
fid = fopen([topic,'.csv'], 'w');

%print the first line for the topic
fprintf(fid,'%s,%s\n',topic,num2str(topicMonthSince2000));

plotx(1)=topicMonthSince2000;
ploty(1)=0;

%axis([xmin xmax ymin ymax])

subplot(6,5,sub);
plot(plotx(1),ploty(1),'--
rs','Marker','v','MarkerEdgeColor','k','MarkerFaceColor','r','MarkerSize',20)
axis([35 80 0 11]);
set(0,'defaultaxesfontsize',6);
xlabel(topic,'FontName','trebuchet','FontSize',10,'FontWeight','bold','Color',[1
0 0]);

hold on;

for i=1:loop
    date=cell2mat(horizon_array(3,i));
    year=str2num(sscanf(date,'%c',4));
    month=str2num(sscanf(date,'%c',6))-year*100;
    day=str2num(sscanf(date,'%c',8))-year*10000-month*100;
    hour=str2num(sscanf(date,'%c',10))-year*1000000-month*10000-day*100;
    hours2000=(((year-2000)*12+month)*30+day)*24+hour;
    MonthSince2000=(((year-2000)*12+month));

    fprintf(fid,
```

```
'%s,%s\n',cell2mat(horizon_array(2,i)),num2str(MonthSince2000));

    plotx(i+1)=MonthSince2000;
    ploty(i+1)=1;
    MarkerSize = 5;

    for j=1:i
        if plotx(j)==MonthSince2000
            ploty(i+1) = ploty(i+1)+1;
        end
    end
    bar(plotx(i+1),ploty(i+1))

end

hold off;
fclose(fid);
```

APPENDIX D – PATH LENGTH IN WIKIPEDIA

By using the "Six degrees of Wikipedia" website (tools.wikimedia.de/sixdeg/)we are able to extract the distance between any two articles in the english edition. The source code for their script is available on their website if you wish to run it in other language edition.

The answer to this web query:

```
http://tools.wikimedia.de/sixdeg/index.jsp?from=Algebra&to=Baseball
```

Is all the HTML code to generate the following page.

Image removed for copyright reasons.
Screen capture of website.

Since we are only interested in the actual number and for all the 870 combinations of the 30 articles being studied. In order to automate this process, we ran the following python script:

Source code:

```
import urllib

list = ("Adenomyosis","Algebra","Aluminium","Baseball","Basketball","Beekeeping","Brigadier",…
"Cellular+automaton","Christmas","Colonization+of+Africa","Color+photography","Criminology",…
"Design","DNA","Elisabeth+of+Bavaria","Entrepreneur","Francisco+Franco","Golf",…
"Hans+Christian+Andersen","History+of+Manchester","Ice+cream","India",…
"Industrial+Revolution","James+Chaney","Locomotive","Massari","Meditation","Moscow",…
"Nobel+Peace+Prize","Paris","Politics","Population","Radio","Stradivarius","World+war+II")

matrix = [[0 for i in range(35)] for i in range(35)]

for k in range(0, 35):
   for j in range (0,35):
      URL = "http://tools.wikimedia.de/sixdeg/index.jsp?from="+list[k]+"&to="+list[j]
      sock = urllib.urlopen(URL)
      htmlSource = sock.read()
      sock.close()
```

```
result = htmlSource.find('degrees of separation</div>')
result = result-2
matrix [k][j] = htmlSource [result]
print "Done: "+list[k]+" "+list[j]
```

APPENDIX E – PATH LENGTH COMPARISONS

Path length between the sample topics in Wikipedia. Length was calculated by using the online tool "Six Degrees of Wikipedia" and automated through a python script (described in Appendix 3)

	Adenomyosis	Algebra	Aluminium	Baseball	Basketball	Beekeeping	Brigadier	Cellular_automaton	Christmas	Colonization_of_Africa	Color_photography	Criminology	Design	DNA	Elisabeth_of_Bavaria	Entrepreneur	Francisco_Franco	Golf	Hans_Christian_Andersen	History_of_Manchester	Ice_cream	India	Industrial_Revolution	James_Chaney	Locomotive	Massari	Meditation	Moscow	Nobel_Peace_Prize	Paris	Politics	Population	Radio	Stradivarius	World_war_II
Adenomyosis	0	4	3	3	4	4	4	4	3	4	4	4	4	3	4	4	3	3	4	4	3	3	3	4	3	5	3	3	3	3	3	3	4	5
Algebra	4	0	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	4	2	1	2	3	3	3	3	2	2	2	2	2	3	4
Aluminium	4	3	0	2	2	3	3	3	2	3	3	3	2	2	3	3	2	2	2	4	2	2	2	3	2	4	3	2	2	2	2	2	4	3
Baseball	4	3	2	0	1	3	3	4	2	3	3	2	3	2	3	3	2	2	3	3	3	2	2	3	2	3	2	2	2	2	2	2	4	3
Basketball	4	2	3	2	0	3	3	4	2	3	3	3	3	3	3	3	2	2	3	4	3	2	2	3	2	4	3	2	2	2	2	2	3	4
Beekeeping	4	3	3	2	2	0	3	3	3	3	3	3	3	2	3	3	3	2	3	4	3	2	2	4	3	4	3	2	2	2	3	2	2	4
Brigadier	4	2	3	3	3	3	0	4	3	3	3	3	3	2	3	3	3	3	3	4	3	2	2	3	3	3	3	2	2	2	3	3	2	4
Cellular_automaton	4	2	3	2	2	3	3	0	3	3	4	3	3	2	3	3	3	2	3	3	3	2	3	3	3	4	3	2	2	2	3	2	2	3
Christmas	3	3	2	2	2	3	3	3	0	3	3	3	3	2	2	3	2	2	1	3	2	2	2	3	2	3	3	2	2	2	2	2	1	4
Colonization_of_Africa	4	2	3	2	2	2	3	4	3	0	3	3	3	3	3	3	2	2	3	4	3	2	2	3	3	3	3	2	2	2	2	2	3	4
Color_photography	4	3	3	3	3	3	4	4	3	3	0	4	3	3	4	3	3	3	3	4	2	3	2	3	2	4	3	3	3	2	3	3	2	4
Criminology	4	2	3	2	2	3	3	4	3	3	4	0	3	2	3	3	3	2	3	4	3	2	2	3	3	4	3	3	3	2	2	2	2	4
Design	4	3	3	3	3	3	4	2	3	4	3	3	0	2	4	3	3	2	4	4	3	2	2	4	3	4	3	3	3	2	1	2	2	4
DNA	3	2	2	2	2	3	3	2	2	3	3	2	3	0	3	3	3	2	3	3	3	2	2	3	2	4	3	2	2	2	3	2	2	4
Elisabeth_of_Bavaria	4	3	3	2	2	3	3	3	2	3	3	3	3	3	0	3	2	2	3	3	3	2	2	3	3	4	3	2	2	2	2	2	4	3
Entrepreneur	4	3	3	2	2	3	4	3	3	3	4	3	3	3	3	0	3	2	3	4	3	2	3	3	3	4	3	2	2	2	2	3	2	3	4
Francisco_Franco	4	3	3	2	2	3	2	3	2	3	3	3	3	2	3	2	0	2	3	3	3	2	2	3	2	3	3	2	2	2	2	2	2	3	4
Golf	4	3	2	2	2	3	3	4	2	3	3	3	3	3	3	3	3	0	3	3	3	2	2	3	3	4	3	2	2	2	2	2	3	4	3
Hans_Christian_Andersen	4	3	3	2	3	3	3	4	1	3	3	3	3	3	3	3	3	2	0	3	3	2	2	2	3	4	3	2	2	2	2	2	2	4	4
History_of_Manchester	4	2	2	2	2	3	2	3	2	3	3	3	2	2	3	2	2	2	3	0	3	2	1	3	2	4	3	2	2	2	2	2	2	3	3
Ice_cream	4	3	3	2	3	3	3	3	2	2	3	3	3	2	2	2	2	2	3	3	0	1	2	3	2	4	3	2	3	2	2	2	2	3	4
India	4	2	2	2	2	3	2	4	2	3	3	3	3	2	3	3	2	2	2	3	2	0	3	2	3	2	3	2	2	2	2	2	2	4	3
Industrial_Revolution	4	3	2	2	2	2	3	3	2	2	2	2	2	2	3	1	2	2	3	3	2	1	0	3	1	4	2	2	2	2	2	2	2	3	3
James_Chaney	4	3	3	2	2	3	3	4	2	3	3	2	3	3	3	3	3	2	3	4	3	2	3	0	3	4	3	2	2	2	2	2	3	3	4
Locomotive	4	3	2	2	2	3	3	4	2	3	3	3	3	2	3	3	2	2	3	3	3	2	2	3	0	4	3	2	2	2	2	2	2	3	3
Massari	4	2	3	2	2	3	3	4	2	3	3	4	3	2	3	3	3	2	3	3	3	2	2	3	2	0	3	2	2	2	2	2	3	3	4
Meditation	3	3	3	2	2	3	3	3	2	3	3	3	3	2	3	3	3	2	3	4	3	2	3	3	3	3	0	2	2	2	3	2	3	4	4
Moscow	4	3	2	2	2	3	3	3	2	3	3	3	2	2	3	2	2	2	3	3	2	1	2	3	2	4	3	0	2	1	2	1	2	3	3
Nobel_Peace_Prize	4	3	3	2	2	3	3	3	2	3	3	3	3	2	3	3	2	2	3	3	3	2	2	2	2	3	2	2	0	2	2	2	2	3	3
Paris	4	2	2	2	1	3	3	4	2	3	3	3	2	2	3	3	2	2	2	3	3	1	2	1	3	2	4	3	1	2	0	2	1	2	3
Politics	4	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	2	2	2	3	4	3	2	2	3	3	4	2	3	3	2	0	2	2	4
Population	3	2	3	2	2	3	3	3	2	3	3	2	3	2	3	3	2	2	3	4	2	1	2	3	3	4	2	2	2	2	2	0	3	3	3
Radio	4	3	2	2	2	2	3	3	2	2	2	3	3	2	3	3	2	2	3	3	2	2	2	2	1	4	3	2	2	2	3	2	0	3	3
Stradivarius	4	3	3	3	2	3	3	4	2	4	3	3	3	3	3	3	2	3	3	4	2	3	3	3	3	4	2	2	3	2	2	2	2	0	4
World_war_II	4	4	4	3	3	4	3	4	3	4	4	4	3	3	4	3	3	3	4	4	3	3	3	3	3	4	4	2	3	2	3	3	3	4	0

Path length between the sample topics in EB print edition. Calculated using UCInet and our EB network for topics

	Adenomyosis	Aluminium	Basketball	Beekeeping	Cellular_automaton	Colonization_of_Africa	Color_photography	Criminology	Design	DNA	Elisabeth_of_Bavaria	Entrepreneur	Francisco_Franco	Golf	Hans_Christian_Andersen	History_of_Manchester	Ice_cream	India	Industrial_Revolution	James_Chaney	Locomotive	Massari	Meditation	Moscow	Nobel_Peace_Prize	Paris	Population	Radio	Stradivarius	World_war_II
Adenomyosis	0	9	11	11	10	10	9	11	9	9	9	7	12	10	13	9	12	9	9	10	12	12	9	9	9	9	11	14	11
Aluminium	9	0	10	10	9	9	8	10	8	4	8	8	11	9	12	8	11	8	8	9	11	11	8	8	8	8	10	13	10
Basketball	11	10	0	11	11	11	10	10	10	10	10	10	11	5	14	8	13	10	10	9	13	13	10	8	8	8	12	15	10
Beekeeping	11	10	11	0	11	11	6	11	8	10	10	9	12	10	14	9	7	10	4	10	9	13	10	9	9	9	8	15	8
Cellular_automaton	10	9	11	11	0	10	9	11	9	5	9	9	12	10	13	9	12	9	9	10	12	12	9	9	9	9	9	11	14	11
Colonization_of_Africa	10	9	11	11	10	0	9	11	9	9	5	9	7	10	13	9	12	5	7	10	12	12	9	9	7	9	9	11	14	5
Color_photography	9	8	10	6	9	9	0	10	2	8	8	8	11	9	8	8	7	8	2	9	7	7	8	8	8	8	8	4	9	6
Criminology	11	10	10	11	11	11	10	0	10	10	10	7	9	9	14	6	13	10	10	5	13	13	10	6	6	6	6	12	15	8
Design	9	8	10	8	9	9	2	10	0	8	8	8	11	9	8	8	9	8	4	9	9	7	8	8	8	8	8	6	9	8
DNA	9	4	10	10	5	9	8	10	8	0	8	8	11	9	12	8	11	8	8	9	11	11	8	8	8	8	8	10	13	10
Elisabeth_of_Bavaria	9	8	10	10	9	5	8	10	8	8	0	8	9	9	12	8	11	4	8	9	11	11	8	8	8	8	8	10	13	8
Entrepreneur	7	8	10	9	9	9	8	7	8	8	8	0	11	9	12	5	11	8	8	8	11	11	8	5	8	5	5	10	13	10
Francisco_Franco	12	11	11	12	12	7	11	9	11	11	9	11	0	10	15	9	14	7	9	10	14	14	11	9	7	9	9	13	16	7
Golf	10	9	5	10	10	10	9	9	9	9	9	9	10	0	13	7	12	9	9	8	12	12	9	7	7	7	7	11	14	9
Hans_Christian_Andersen	13	12	14	14	13	13	8	14	8	12	12	12	15	13	0	12	15	10	12	13	15	11	12	12	12	12	12	13	14
History_of_Manchester	9	8	8	9	9	9	8	6	8	8	8	5	9	7	12	0	11	8	8	5	11	11	8	2	6	2	2	10	13	8
Ice_cream	12	11	13	7	12	12	7	13	9	11	11	11	14	12	15	11	0	11	5	12	10	14	11	11	11	11	11	9	16	9
India	9	8	10	10	9	5	8	10	8	8	4	8	7	9	10	8	11	0	6	9	11	11	6	8	6	8	8	10	13	4
Industrial_Revolution	9	8	10	4	9	7	2	10	4	8	8	8	9	9	12	8	5	6	0	9	5	9	8	8	6	8	8	4	11	4
James_Chaney	10	9	9	10	10	10	9	5	9	9	9	8	10	8	13	5	12	9	9	0	12	12	9	5	7	5	5	11	14	9
Locomotive	12	11	13	9	12	12	7	13	9	11	11	11	14	12	15	11	10	11	5	12	0	14	11	11	11	11	11	9	16	9
Massari	12	11	13	13	12	12	7	13	7	11	11	11	14	12	11	11	14	11	9	12	14	0	11	11	11	11	11	11	10	13
Meditation	9	8	10	10	9	9	8	10	8	8	8	8	11	9	12	8	11	6	8	9	11	11	0	8	8	8	8	10	13	10
Moscow	9	8	8	9	9	9	8	6	8	8	8	5	9	7	12	2	11	8	8	5	11	11	8	0	6	2	2	10	13	8
Nobel_Peace_Prize	9	8	8	9	9	7	8	6	8	8	8	8	7	7	12	6	11	6	6	7	11	11	8	6	0	6	6	10	13	4
Paris	9	8	8	9	9	9	8	6	8	8	8	5	9	7	12	2	11	8	8	5	11	11	8	2	6	0	2	10	13	8
Population	9	8	8	9	9	9	8	6	8	8	8	5	9	7	12	2	11	8	8	5	11	11	8	2	6	2	0	10	13	8
Radio	11	10	12	8	11	11	4	12	6	10	10	10	13	11	12	10	9	10	4	11	9	11	10	10	10	10	10	0	13	8
Stradivarius	14	13	15	15	14	14	9	15	9	13	13	13	16	14	13	13	16	13	11	14	16	10	13	13	13	13	13	13	0	15
World_war_II	11	10	10	8	11	5	6	8	8	10	8	10	7	9	14	8	9	4	4	9	9	13	10	8	4	8	8	8	15	0

APPENDIX F – CIRCLE OF KNOWLEDGE STRUCTURE

A "1" denotes a link between the two pairs. From this structure is the Circle of Knowledge represented in a .vna file and then imported into UCINET. The UCINET equivalent structure requires a 229 by 229 matrix that is difficult to display and understand in a document like this. Therefore, this is the display of the .vna file contents.

*tie data

To	from	links
The_Circle_of_Knowledge	Part_One:Matter_and_Energy	1
The_Circle_of_Knowledge	Part_Two:The_Earth	1
The_Circle_of_Knowledge	Part_Three:Life_on_Earth	1
The_Circle_of_Knowledge	Part_Four:Human_Life	1
The_Circle_of_Knowledge	Part_Five:Human_Society	1
The_Circle_of_Knowledge	Part_Six:Art	1
The_Circle_of_Knowledge	Part_Seven:Technology	1
The_Circle_of_Knowledge	Part_Eight:Religion	1
The_Circle_of_Knowledge	Part_Nine:The_History_of_Mankind	1
The_Circle_of_Knowledge	Part_Ten:The_Branches_of_Knowledge	1
Part_One:Matter_and_Energy	Division_I:Atoms:Atomic_Nuclei_and_Elementary_Particles	1
Part_One:Matter_and_Energy	Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	1
Part_One:Matter_and_Energy	Division_III:The_Universe:Galaxies__Stars__the_Solar_System	1
Division_I:Atoms:Atomic_Nuclei_and_Elementary_Particles	The_Structure_and_Properties_of_Atoms	1
Division_I:Atoms:Atomic_Nuclei_and_Elementary_Particles	The_Atomic_Nucleus_and_Elementary_Particles	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Chemical_Elements:Periodic_Variation_in_Their_Properties	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Chemical_Compounds:Molecular_Structure_and_Chemical_Bonding	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Chemical_Reactions	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Heat__Thermodynamics__and_the_Nonsolid_States_of_Matter	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	The_Solid_State_of_Matter	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Mechanics_of_Particles__Rigid_Bodies__and_Deformable_Bodies:Elasticity__Vibrations__and_Flow	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Electricity_and_Magnetism	1
Division_II:Energy__Radiation__and_the_States_and_Transformation_of_Matter	Waves_and_Wave_Motion	1

Division_III:The_Universe:Galaxies__Stars__the_Solar_System	The_Cosmos	1
Division_III:The_Universe:Galaxies__Stars__the_Solar_System	Galaxies_and_Stars	1
Division_III:The_Universe:Galaxies__Stars__the_Solar_System	The_Solar_System	1
Part_Two:The_Earth	Division_I:The_Earth's_Properties__Structure__and_Composition	1
Part_Two:The_Earth	Division_II:The_Earth's_Envelope:Its_Atmosphere_and_Hydrosphere	1
Part_Two:The_Earth	Division_III:The_Earth's_Surface_Features	1
Part_Two:The_Earth	Division_IV:The_Earth's_History	1
Division_I:The_Earth's_Properties__Structure__and_Composition	The_Planet_Earth	1
Division_I:The_Earth's_Properties__Structure__and_Composition	The_Earth's_Physical_Properties	1
Division_I:The_Earth's_Properties__Structure__and_Composition	The_Structure_and_Composition_of_the_Earth's_Interior	1
Division_I:The_Earth's_Properties__Structure__and_Composition	The_Earth's_Constituent_Minerals_and_Rocks	1
Division_II:The_Earth's_Envelope:Its_Atmosphere_and_Hydrosphere	The_Atmosphere	1
Division_II:The_Earth's_Envelope:Its_Atmosphere_and_Hydrosphere	The_Hydrosphere:the_Oceans__Freshwater_Bodies__and_Ice_Masses	1
Division_II:The_Earth's_Envelope:Its_Atmosphere_and_Hydrosphere	Weather_and_Climate	1
Division_III:The_Earth's_Surface_Features	Physical_Features_of_the_Earth's_Surface	1
Division_III:The_Earth's_Surface_Features	Features_Produced_by_Geomorphic_Processes_Acting_on_the_Earth's_Surface	1
Division_IV:The_Earth's_History	Origin_and_Development_of_the_Earth_and_Its_Envelopes	1
Division_IV:The_Earth's_History	The_Interpretation_of_the_Geological_Record	1
Division_IV:The_Earth's_History	The_Eras_and_Periods_of_Geological_Time	1
Part_Three:Life_on_Earth	Division_I:The_Nature_and_Diversity_of_Living_Things	1
Part_Three:Life_on_Earth	Division_II:The_Molecular_Basis_of_Vital_Processes	1
Part_Three:Life_on_Earth	Division_III:The_Structures_and_Functions_of_Organisms	1
Part_Three:Life_on_Earth	Division_IV:Behavorial_Responses_of_Organisms	1
Part_Three:Life_on_Earth	Division_V:The_Biosphere:the_World_of_Living_Things	1
Division_I:The_Nature_and_Diversity_of_Living_Things	Characteristics_of_Living_Things	1
Division_I:The_Nature_and_Diversity_of_Living_Things	The_Origin_of_Life_and_the_Evolution_of_Living_Things	1
Division_I:The_Nature_and_Diversity_of_Living_Things	The_Classification_of_Living_Things	1
Division_II:The_Molecular_Basis_of_Vital_Processes	Chemicals_and_the_Vital_Processes	1
Division_II:The_Molecular_Basis_of_Vital_Processes	Metabolism:Bioenergetics_and_Biosynthesis	1
Division_II:The_Molecular_Basis_of_Vital_Processes	Vital_Processes_at_the_Molecular_Level	1

Division_III:The_Structures_and_Functions_of_Organisms	The_Cellular_Basis_of_Form_and_Function	1
Division_III:The_Structures_and_Functions_of_Organisms	The_Relation_of_Form_and_Function_in_Organisms	1
Division_III:The_Structures_and_Functions_of_Organisms	Coordination_of_Vital_Processes:Regulation_and_Integration	1
Division_III:The_Structures_and_Functions_of_Organisms	Covering_and_Support:Integumentary__Skeletal__and_Musculatory_Systems	1
Division_III:The_Structures_and_Functions_of_Organisms	Nutrition:the_Procurement_and_Processing_of_Nutrients	1
Division_III:The_Structures_and_Functions_of_Organisms	Gas_Exchange__Internal_Transport__and_Elimination	1
Division_III:The_Structures_and_Functions_of_Organisms	Reproduction_and_Sex	1
Division_III:The_Structures_and_Functions_of_Organisms	Development:Growth__Differentiation__and_Morphogenesis	1
Division_III:The_Structures_and_Functions_of_Organisms	Heredity:the_Transmission_of_Traits	1
Division_IV:Behavorial_Responses_of_Organisms	Nature_and_Patterns_of_Behavorial_Responses	1
Division_IV:Behavorial_Responses_of_Organisms	Development_and_Range_of_Behavioral_Capacities:Individual_and_Group_Behaviour	1
Division_V:The_Biosphere:the_World_of_Living_Things	Basic_Features_of_the_Biosphere	1
Division_V:The_Biosphere:the_World_of_Living_Things	Biological_Populations_and_Communities	1
Division_V:The_Biosphere:the_World_of_Living_Things	Hazards_of_Life_in_the_Biosphere:Disease_and_Death	1
Division_V:The_Biosphere:the_World_of_Living_Things	Biogeographic_Distribution_of_Organisms:Ecosystems	1
Division_V:The_Biosphere:the_World_of_Living_Things	The_Place_of_Humans_in_the_Biosphere	1
Part_Four:Human_Life	Division_I:Stages_in_the_Development_of_Human_Life_on_Earth	1
Part_Four:Human_Life	Division_II:The_Human_Organism:Health_and_Disease	1
Part_Four:Human_Life	Division_III:Human_Behaviour_and_Experience	1
Division_I:Stages_in_the_Development_of_Human_Life_on_Earth	Human_Evolution	1
Division_I:Stages_in_the_Development_of_Human_Life_on_Earth	Human_Heredity:the_Races_of_Mankind	1
Division_II:The_Human_Organism:Health_and_Disease	The_Structures_and_Functions_of_the_Human_Body	1
Division_II:The_Human_Organism:Health_and_Disease	Human_Health	1
Division_II:The_Human_Organism:Health_and_Disease	Human_Diseases	1
Division_II:The_Human_Organism:Health_and_Disease	The_Practice_of_Medicine_and_the_Care_of_Health	1
Division_III:Human_Behaviour_and_Experience	Human_Nature_and_Experience:General_Considerations	1
Division_III:Human_Behaviour_and_Experience	Influence_of_the_Current_Environment_on_a_Person's_Behaviour_and_Conscious_Experience:Attention__Sensation__and_Perception	1
Division_III:Human_Behaviour_and_Experience	Current_Internal_States_Affecting_a_Person's_Behaviour_and_Conscious_Experience	1

Division_III:Human_Behaviour_and_Experience	Persisting_Capacities_and_Inclinations_That_Influence_Human_Behaviour_and_Conscious_Experience	1
Division_III:Human_Behaviour_and_Experience	Development_of_a_Person's_Potentials:Learning_and_Thinking	1
Division_III:Human_Behaviour_and_Experience	Personality_and_the_Self:Integration_and_Disintegration_of_the_Person_as_a_Whole	1
Part_Five:Human_Society	Division_I:Social_Groups:Peoples_and_Cultures	1
Part_Five:Human_Society	Division_II:Social_Organization_and_Social_Change	1
Part_Five:Human_Society	Division_III:The_Production__Distribution__and_Utilization_of_Wealth	1
Part_Five:Human_Society	Division_IV:Politics_and_Government	1
Part_Five:Human_Society	Division_V:Law	1
Part_Five:Human_Society	Division_VI:Education	1
Division_I:Social_Groups:Peoples_and_Cultures	Peoples_and_Cultures_of_the_World	1
Division_I:Social_Groups:Peoples_and_Cultures	The_Development_of_Human_Culture	1
Division_I:Social_Groups:Peoples_and_Cultures	Major_Cultural_Components_and_Institutions_of_Human_Societies	1
Division_I:Social_Groups:Peoples_and_Cultures	Language_and_Communication	1
Division_II:Social_Organization_and_Social_Change	Social_Structure_and_Change	1
Division_II:Social_Organization_and_Social_Change	The_Group_Structure_of_Society	1
Division_II:Social_Organization_and_Social_Change	Social_Status	1
Division_II:Social_Organization_and_Social_Change	Human_Populations:Urban_and_Rural_Communities	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	Economic_Concepts__Issues__and_Systems	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	The_Consumer_and_the_Market:Pricing_and_the_Mechanisms_for_Distributing_Goods	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	The_Organization_of_Production_and_Distribution	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	The_Distribution_of_Income_and_Wealth	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	Macroeconomics	1
Division_III:The_Production__Distribution__and_Utilization_of_Wealth	Economic_Growth_and_Planning	1
Division_IV:Politics_and_Government	Political_Theory	1
Division_IV:Politics_and_Government	Political_Institutions:the_Structure__Branches__and_Offices_of_Government	1
Division_IV:Politics_and_Government	The_Functioning_of_Government:the_Dynamics_of_the_Political_Process	1
Division_IV:Politics_and_Government	International_Relations:Peace_and_War	1
Division_V:Law	Philosophies_and_Systems_of_Law__the_Practice_of_Law	1
Division_V:Law	Branches_of_Public_Law__Substantive_and_Procedural	1
Division_V:Law	Branches_of_Private_Law__Substantive_and_Procedural	1

Division_VI:Education	The_Aims_and_Organization_of_Education	1
Division_VI:Education	Education_Around_the_World	1
Part_Six:Art	Division_I:Art_in_General	1
Part_Six:Art	Division_II:The_Particular_Arts	1
Division_I:Art_in_General	Theory_and_Classification_of_the_Arts	1
Division_I:Art_in_General	Experience_and_Criticism_of_Works_of_Art__the_Nonaesthetic_Context_of_Art	1
Division_I:Art_in_General	Characteristics_of_the_Arts_in_Particular_Cultures	1
Division_II:The_Particular_Arts	Literature	1
Division_II:The_Particular_Arts	Theatre	1
Division_II:The_Particular_Arts	Motion_Pictures	1
Division_II:The_Particular_Arts	Music	1
Division_II:The_Particular_Arts	Dance	1
Division_II:The_Particular_Arts	Architecture__Garden_and_Landscape_Design__and_Urban_Design	1
Division_II:The_Particular_Arts	Sculpture	1
Division_II:The_Particular_Arts	Drawing__Painting__Printmaking__and_Photography	1
Division_II:The_Particular_Arts	Arts_of_Decoration_and_Functional_Design	1
Part_Seven:Technology	Division_I:The_Nature_and_Development_of_Technology	1
Part_Seven:Technology	Division_II:Elements_of_Technology	1
Part_Seven:Technology	Division_III:Major_Fields_of_Technology	1
Division_I:The_Nature_and_Development_of_Technology	Technology:Its_Scope_and_History	1
Division_I:The_Nature_and_Development_of_Technology	The_Organization_of_Human_Work	1
Division_II:Elements_of_Technology	Technology_of_Energy_Conversion_and_Utilization	1
Division_II:Elements_of_Technology	Technology_of_Tools_and_Machines	1
Division_II:Elements_of_Technology	Technology_of_Measurement__Observation__and_Control	1
Division_II:Elements_of_Technology	Extraction_and_Conversion_of_Industrial_Raw_Materials	1
Division_II:Elements_of_Technology	Technology_of_Industrial_Production_Processes	1
Division_III:Major_Fields_of_Technology	Agriculture_and_Food_Production	1
Division_III:Major_Fields_of_Technology	Technology_of_the_Major_Industries	1
Division_III:Major_Fields_of_Technology	Construction_Technology	1
Division_III:Major_Fields_of_Technology	Transportation_Technology	1
Division_III:Major_Fields_of_Technology	Technology_of_Information_Processing_and_of_Communications_Systems	1
Division_III:Major_Fields_of_Technology	Military_Technology	1
Division_III:Major_Fields_of_Technology	Technology_of_the_Urban_Community	1
Division_III:Major_Fields_of_Technology	Technology_of_Earth_and_Space_Exploration	1
Part_Eight:Religion	Division_I:Religion_in_General	1
Part_Eight:Religion	Division_II:The_Particular_Religions	1
Division_I:Religion_in_General	Knowledge_and_Understanding_of_Religion	1

Division_I:Religion_in_General	The_Religious_Life:Institutions_and_Practices	1
Division_II:The_Particular_Religions	Prehistoric_Religion_and_Primitive_Religion	1
Division_II:The_Particular_Religions	Religions_of_Ancient_Peoples	1
Division_II:The_Particular_Religions	Hinduism_and_Other_Religions_of_India	1
Division_II:The_Particular_Religions	Buddhism	1
Division_II:The_Particular_Religions	Indigenous_Religions_of_East_Asia:Religions_of_China__Korea__and_Japan	1
Division_II:The_Particular_Religions	Judaism	1
Division_II:The_Particular_Religions	Christianity	1
Division_II:The_Particular_Religions	Islam	1
Division_II:The_Particular_Religions	Other_Religions_and_Religious_Movements_in_the_Modern_World	1
Part_Nine:The_History_of_Mankind	Division_I:Peoples_and_Civilizations_of_Ancient_Southwest_Asia__North_Africa__and_Europe	1
Part_Nine:The_History_of_Mankind	Division_II:Peoples_and_Civilizations_of_Medieval_Europe__North_Africa__and_Southwest_Asia	1
Part_Nine:The_History_of_Mankind	Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	1
Part_Nine:The_History_of_Mankind	Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	1
Part_Nine:The_History_of_Mankind	Division_V:Peoples_and_Civilizations_of_Pre-Columbian_America	1
Part_Nine:The_History_of_Mankind	Division_VI:The_Modern_World_to_1920_	1
Part_Nine:The_History_of_Mankind	Division_VII:The_World_Since_1920_	1
Division_I:Peoples_and_Civilizations_of_Ancient_Southwest_Asia__North_Africa__and_Europe	Early_Peoples_and_Civilizations_of_Southwest_Asia_and_Egypt__the_Aegean__and_North_Africa	1
Division_I:Peoples_and_Civilizations_of_Ancient_Southwest_Asia__North_Africa__and_Europe	Peoples_of_Ancient_Europe_and_the_Classical_Civilizations_of_the_Ancient_Mediterranean_World_to_AD_395_	1
Division_II:Peoples_and_Civilizations_of_Medieval_Europe__North_Africa__and_Southwest_Asia	Western_Europe__the_Byzantine(Eastern_Roman)_Empire__and_Eastern_Europe_from_AD_395_to_c._1050_	1
Division_II:Peoples_and_Civilizations_of_Medieval_Europe__North_Africa__and_Southwest_Asia	The_Formative_Period_in_Islamic_History__from_AD_622_to_c._1055_	1
Division_II:Peoples_and_Civilizations_of_Medieval_Europe__North_Africa__and_Southwest_Asia	Western_Christendom_in_the_High_and_Later_Middle_Ages_(c._1050_-c._1500)	1
Division_II:Peoples_and_Civilizations_of_Medieval_Europe__North_Africa__and_Southwest_Asia	The_Crusading_Movement__the_Islamic_States_of_Southwest_Asia__North_Africa__and_Europe__and_the_States_of_Eastern_Christendom_from_c._1050_to_c._1480_	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	China_to_the_Beginning_of_the_Late_T'ang(AD_755)	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	China_from_the_Late_T'ang_(AD_755)_to_the_Late_Ch'ing(c.1839)	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	Inner_(Central_and_Northeast)_Asia_to_c._1750_	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	Japan_to_the_Meiji_Restoration_(1868)__and_Korea_to_1910_	1

Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	The_Indian_Subcontinent_and_Ceylon_to_c._AD_1200_	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	The_Indian_Subcontinent_from_c._1200_to_1761__and_Ceylon_from_c._1200_to_1505_	1
Disision_III:Peoples_and_Traditional_Civilizations_of_East__Central__South__and_Southeast_Asia	The_Peoples_and_Civilizations_of_Southeast_Asia_to_c._1600_	1
Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	West_Africa_to_c._1885_	1
Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	The_Nilotic_sudan_and_Ethiopia_from_c._AD_550_to_1885_	1
Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	East_Africa_and_Madagascar_to_c._1885_	1
Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	Central_Africa_to_c._1885_	1
Division_IV:Peoples_and_Civilizations_of_Sub-Saharan_Africa_to_1885_	Southern_Africa_to_c._1885_	1
Division_V:Peoples_and_Civilizations_of_Pre-Columbian_America	Andean_Civilization_to_c._AD_1540_	1
Division_V:Peoples_and_Civilizations_of_Pre-Columbian_America	Meso-American_Civilization_to_c._AD_1540_	1
Division_VI:The_Modern_World_to_1920_	Western_Europe_from_c._1500_to_c._1789_	1
Division_VI:The_Modern_World_to_1920_	Eastern_Europe__Southwest_Asia__and_North_Africa_from_c._1480_to_c._1800_	1
Division_VI:The_Modern_World_to_1920_	Europe_from_1789_to_c._1920_	1
Division_VI:The_Modern_World_to_1920_	European_Colonies_in_the_Americas_from_1492_to_c._1790_	1
Division_VI:The_Modern_World_to_1920_	Development_of_the_United_States_and_Canada_from_1763_to_1920_	1
Division_VI:The_Modern_World_to_1920_	Development_of_the_Latin-American_and_Caribbean_Nations_to_c._1920_	1
Division_VI:The_Modern_World_to_1920_	Australia_and_Oceania_to_c._1920_	1
Division_VI:The_Modern_World_to_1920_	South_Asia_Under_the_Influence_of_European_Imperialism_from_c._1500_to_c._1920_	1
Division_VI:The_Modern_World_to_1920_	Southeast_Asia_Under_the_Influence_of_European_Imperialism_to_c._1920_	1
Division_VI:The_Modern_World_to_1920_	China_from_1839_Until_the_Onset_of_Revolution_(to_c._1911)__and_Japan_from_the_Meiji_Restoration_to_c._1910_	1
Division_VI:The_Modern_World_to_1920_	Southwest_Asia_and_North_Africa_(c._1800-1920)__and_Sub-Saharan_Africa_(1885-c._1920)_Under_the_Influence_of_European_Imperialism:_the_Early_Colonial_Period	1
Division_VII:The_World_Since_1920_	International_Movements__Diplomacy__and_War_Since_1920_	1
Division_VII:The_World_Since_1920_	Europe_Since_c._1920_	1
Division_VII:The_World_Since_1920_	the_United_States_and_Canada_Since_1920_	1
Division_VII:The_World_Since_1920_	Latin-American_and_Caribbean_Nations_since_c._1920_	1
Division_VII:The_World_Since_1920_	East_Asia:_China_in_Revolution__the_Era_of_Jap	1

	anese_Hegemony__and_the_Influence_of_the_United_States_in_the_20th_Century	
Division_VII:The_World_Since_1920_	South_and_Southeast_Asia:_the_Late_Colonial_Period_and_the_Emergence_of_New_Nations_Since_1920_	1
Division_VII:The_World_Since_1920_	Australia_and_Oceania_Since_1920_	1
Division_VII:The_World_Since_1920_	Southwest_Asia_and_Africa:_the_Late_Colonial_Period_and_the_Emergence_of_New_Nations_in_the_20th_Century	1
Part_Ten:The_Branches_of_Knowledge	Division_I:_Logic	1
Part_Ten:The_Branches_of_Knowledge	Division_II:_Mathematics	1
Part_Ten:The_Branches_of_Knowledge	Division_III:_Science	1
Part_Ten:The_Branches_of_Knowledge	Division_IV:_History_and_the_Humanities	1
Part_Ten:The_Branches_of_Knowledge	Division_V:_Philosophy	1
Part_Ten:The_Branches_of_Knowledge	Division_VI:_Preservation_of_Knowledge	1
Division_I:_Logic	History_and_Philosophy_of_Logic	1
Division_I:_Logic	Formal_Logic__Metalogic__and_Applied_Logic	1
Division_II:_Mathematics	History_and_Foundations_of_Mathematics	1
Division_II:_Mathematics	Branches_of_Mathematics	1
Division_II:_Mathematics	Applications_of_Mathematics	1
Division_III:_Science	History_and_Philosophy_of_Science	1
Division_III:_Science	The_Physical_Sciences	1
Division_III:_Science	The_Earth_Sciences	1
Division_III:_Science	The_Biological_Sciences	1
Division_III:_Science	Medicine_and_Affiliated_Disciplines	1
Division_III:_Science	The_Social_Sciences_and_Psychology_and_Linguistics	1
Division_III:_Science	The_Technological_Sciences	1
Division_IV:_History_and_the_Humanities	Historiography_and_the_Study_of_History	1
Division_IV:_History_and_the_Humanities	The_Humanities_and_Humanistic_Scholarship	1
Division_V:_Philosophy	History_of_Philosophy	1
Division_V:_Philosophy	The_Nature_and_the_Divisions_of_Philosophy	1
Division_V:_Philosophy	Philosophical_Schools_and_Doctrines	1
Division_VI:_Preservation_of_Knowledge	Institutions_and_Techniques_for_the_Collection__Storage__Dissemination__and_Preservation_of_Knowledge	1