Running head: WHAT TELETEX CAN TEACH US

What Teletex can teach us

Jay McKinnon

Simon Fraser University, Burnaby BC Canada

School of Communication

03 June 2011

Abstract

Through much of the 1980s, governments and corporations invested hundreds of millions of hours and dollars developing several technologies now generally considered extinct: teletext, videotex, Telidon and Minitel. The technologies which now define the history of the Internet were relatively ignored in their time, though known to exist as obscure academic experiments. Most curiously, BBSes and Fidonet – popular technologies which defined the digital world for millions – were, and remain, largely ignored by scholars.

This paper quantitatively explores the scholarly output regarding ICT paradigms extant in the 1980s, reports on the current state of development of the relevant technologies and theoretically interrogates the institutional lessons to which can be learned from failed research programs.

Table of Contents

List of Figures & Tables

What Teletex can teach us

> Since there is no clear line available to demarcate the *modem* from the *postmodem*, the latter comes to express both an 'end' of the former and a radical continuation of it.[1]

Introduction

Through much of the 1980s, governments and corporations invested hundreds of millions of hours and dollars developing three technologies now generally considered extinct: teletext, videotex and Minitel. At that time, the technologies which now define the accepted history of the Internet were known to exist, though only as obscure academic experiments. Most curiously, BBSes and Fidonet – popular technologies which defined the cyberspace for millions – were, and remain, largely overlooked by scholars. This paper quantitatively explores the scholarly output regarding ICT paradigms extant in the 1980s, reports on the current state of development of the relevant technologies and theoretically interrogates the institutional lessons to which can be learned from failed research programs. Keyword frequency analysis of different databases are used to timeline of scholarly interest before drawing larger historical lessons from select documents.

Still exploratory in nature, this study was inspired by the following contradiction: anecdotal evidence suggests that the vast majority of early-adopters of the Internet were

---

[1] This quote took on new meaning after passing through the GoogleBooks optical character recognition (OCR) system. The original appears in a chapter, by Michael Power, entitled "Modernism, postmodernism and organization" (Hassard & Pym, 1990).

frequent users of Bulletin Board Systems (BBSes), yet the influence of the BBS phenomenon, in configuring the home user for computer communication, is largely absent from scholarship on the reception of the Internet. Historical investigation of the state of digital communication research in the 1970s and 1980s revealed other technologies which, now rarely mentioned, commanded a great deal of scholarly attention. This observation raises further questions of what influence those forgotten studies, especially by graduate students, had on the rapid adoption of Internet-related topics by academics in the 1990s.

Well over a hundred years ago, the epithet “Whig History” was coined to criticise scholars who represented history as a coherent, teleological march of progress toward the inevitable (morally superior) present. In the later half of the 20th century, social scientists like Thomas Kuhn, Trevor Pinch, Wiebe Bijker and Bruno Latour demonstrated the messy, irregular, political, and often irrational processes by which science is really practiced and new technologies developed. Even while rationalizing the process of adoption and diffusion, Roger Everett was struck by the extent to which end-user agency could overthrow the designs of engineers. In the face of this overwhelming body of scholarship demanding nuance and complexity in the history of technology, the general consensus on the historical origins of today’s Internet is truly remarkable.

American movie fans might recognize the generally accepted history of Internet as a “Field of Dreams” hypothesis: they built it and we came. After 30 or 40 years of development in university laboratories, the infrastructure was ready for Tim Berners-Lee to invent the World Wide Web at CERN, and then everybody rushed to download

Netscape and start a Geocities webpage. Nowhere in this narrative arch is the process by which tens of millions of end-users developed the skills necessary to use a modem with a personal computer or, more importantly, developed the conceptual frame necessary to conceive of what the Internet might be and how they might use it. The rapid adoption of the Internet is ill-explained by a model of technological change in which an invention has a single-point of origin and is propagated by early-adopters and change agents.

If, instead, we take seriously the theorists mentioned, then we should look not for evidence of adoption of successful technologies, or the abandonment of failed technologies, but for what Everett called the “replacement discontinuance”: the shift from one technology to another. It is in this process that stakeholders convert their metaphorical frames from one technology to another, often recreating a familiar pattern of practice in the new medium.

In the rise and fall of the keyword clusters, we observe the adoption and discontinuance of technological research programs by students, academics and technical intellectuals. When the curves are combined, we may even find directions to the moments of social conflict and technical plasticity predicted by Kuhn, Pinch, Bijker, Latour and Everett.

Method: keyword frequency analysis

Defining the technological programs

For the purpose of this study. six research programs are treated as distinct technologies defined by clusters of keywords. They are grouped under the headings of mainframe systems, teletext, videotex, national projects, bulletin board systems, and the Internet. Mainframe systems, generally time-share computing services descendent from the ENIAC, AT&T and IBM systems, are widely recognized in North America by corporate brands like AOL, Compuserv, Prodigy, The Source, the Well, and in academic systems like PLATO. They, like the Internet, which needs no introduction, are present in this study only for comparative purposes.

Teletext and videotex were sibling technologies developed out of the technical ability to embed digital information in the vertical blanking interval of television signals. These two exhibit similar models of corporate and institutional sponsorship for research and differ primarily in the means of transmission: teletext via broadcast television signals and videotex via telephone lines or packet switched networks. In their heyday, boosters of these technologies – generally from either the television or newspaper industries – travelled in the same circles and traded barbs in the same forums. The two national projects examined in this study – Minitel and Telidon – were similar to videotex (though both had teletext components) but enjoyed a higher level of state-sponsorship which resulted in unique development patterns.

Bulletin Board Systems were a home computer-based technology, which distributed of digital information across public telephone networks. While BBS technologies were adopted by governments and corporations after reaching maturity, they are unique among the research programs in that the primary drivers of innovation and adoption were hobbyists and home users.

Data sources and keyword clusters

The two data sources sampled in this study are: (1) the WorldCat.org database of theses and dissertations published; and (2) Google Ngram database of books published in English. Keywords were selected, clustered and tested as markers of six distinct research programs: mainframe systems, teletext, videotex, national project, bulletin board systems, and the Internet.

The Teletext Cluster includes the terms CEEFAX, the first such system developed by the BBC, teletext and its less-popular synonym[2], videotext. The Videotex Cluster includes the terms Viewdata, developed by the British Post Office, the PRESTEL branded-service, videotex and and its less-popular synonym teletex. The National Projects cluster includes the terms Telidon, from Canada, and Minitel, from France (but not Antipode). The Bulletin Board System Cluster includes two terms: BBS and FidoNet. The BBS acroynm exhibits low frequency low-frequency prior to the development of the technology because of its use in different fields but a manual review of the source records confirms that the keyword's growth is primarily attributable to the ICT. FidoNet was, and

[2] Gary Arlen's submission to the conference proceedings of Videotex ’80, contains a humorous three-page subsection entitled "Pitfalls in the Name Game" (Arlen 1980, p.7-9). Therein, he discusses the frustrating drift and mutation between teletext/teletex/videotext/videotex, since the terms were first defined by the International Consultative Committee on Telegraph and Telephone (CITT) in 1978. While the etymological controversy is an interesting case study for ANT and SCOT scholars, it is outside the scope of this paper. While the last letter distinguished the two technologies more often than the first, general practice appears to have eventually settled on “teletext” and “videotext”.

remains, the largest network of BBSes. The terms "bulletin board system", "bulletin board service" and "electronic bulletin board" could not be queried because the 3gram database is too large to process on the available equipment. Unfortunately, the many of most popular mainframe systems, like PLATO, Prodigy, The Source and The Well, exibit low-uniqueness in their names so the terms had to be discarded, leaving only CompuServ (a branded coinage) and AOL in the Mainframe Cluster. Finally, the Internet Cluster includes the terms ARPAnet, USENET, UUNet, HTTP, WWW and “Internet”.

When keyword frequencies are reported as percentages, they are percentages of the total corpus for that year so that the increasing rate of publication or digitization does not artificically inflate measurements.

Sources of error and bias

Any attempt to perform quantitative textual analysis on historical records encounters several of sources of error, some obvious and others hidden. Databases are always incomplete, biased towards one language or another, and often over-represent some repositories while omitting others. In this study, the bias of omission is most noticeable when searching for trade publications known to have existed: LexisNexus contains few and GoogleBooks’ Ngrams contains none. Conference proceedings and popular computer magazines (like Byte Magazine and InfoWeek) were examined as part of the qualitative investigation, but are not represented in the quantitative results.

Furthermore, when pre-digital texts are scanned, the optical character recognition (OCR) process introduces new errors which may inflate, or deflate, keyword counts; if OCR errors mutate a more common word into a keyword it may necessitate the elimination of an important sample or, worse, lead to erroneous conclusions. The opening quote to this paper illustrates one such OCR error discovered in the mutation of “modern” into “modem”, the latter being the piece of equipment used to connect a personal computer to a telephone line, and former being one of the most popular words in Anglophone academia. The term “Internet” exhibits similar problems in OCR’d database records, which results in a great many mentions dating back to the 15th century. No effort was made to eliminate those erroneous hits because “Internet” becomes so popular during the period under study that it overwhelms both intra-keyword error and inter-keyword comparisons: since the year 2000, over 50% of books contained the term “Internet” and over 70% contained a webpage URL. As mentioned above, some chosen keywords (especially company names) were eliminated because they had other, more common usages.

In the absence of the ability to use concordances to verify keyword hits, this study occasionally takes the unorthodox approach of reporting the sum for each cluster as if each keyword hit were a distinct text. For example, a book in the Ngram database which mentions Viewdata, PRESTEL, videotex and teletext will be reported four times for the cluster, inflating the true frequency. This approach undermines statistical comparisons between clusters in favour of greater confidence that the differences observed within clusters are real and not the result of typos or OCR mutations. Table 3 reports the composition of the keyword clusters of primary interest to this study – videotex, teletext, national projects, mainframes and BBSes – as a percent of books in the Ngram database for each year.

Counting hits: quantitative analysis of textual databases

In “Quantitative Analysis of Culture Using Millions of Digitized Books” (Michel et al, 2010), and the related supplement, the authors of the GoogleBooks Ngram database demonstrate how keyword frequencies can reflect the adoption of technology and other cultural phenomena. In contrast to that study, which tracked 154 successful inventions from the 17th century, this study tracks a smaller number of “failed” technologies over just the last four decades.

Similar work has used citations counts in academic databases like Web of Science, Scopus or, increasingly, Google Scholar, to measure the influence of an article or scholar. Concerns about transparency of the search algorithms and the “low overlap or high uniqueness between the three tools” has contributed to the consensus that, when doing citation counts, Google Scholar should be used in conjunction with the other two major citation tracking engines. Google Scholar hits on specified keywords are now widely accepted as a crude measure of interest in a topic of study: 15,000 hits for “Sesame Street” generally indicates that the children’s program has attracted more scholarly interest than “Teletubbies”, which has 2,300 hits. While this study does not use GoogleScholar queries (many of the data gathering tools are explicitly prohibited by the Terms of Service agreement), much of the scholarship and criticism of GoogleScholar applies to the use of frequency analysis on the Ngram and WorldCat databases.

Queries of the googlebooks-eng-all-20090715 Ngram database provide keyword frequencies for English-language books. Filters provided by the WorldCat.org’s database of library holdings provide counts of theses and dissertations tagged with the same keywords. Not withstanding concerns about the impact of search engines on the production of academic knowledge (van Dijck, 2010), these keyword counts suggest provocative questions about the development of information communication technologies (ICTs) in the 20th century,

Research programs as reflected in theses and dissertations (WorldCat)

Table 1 lists, by year, the number of theses and dissertations in each technology cluster, as reported by the Worldcat database. By terminating this graph in 1994, before the “dot-com boom”, we can represent the Internet Cluster with linear-scaled axes, as in Figure 1, a stacked column graph. In Figure 2, which excludes the Internet Cluster, it is evident that Social Media cluster keywords have only recently begun to feature in graduate work, while videotex, teletext and national projects were in their heyday in the 1980s. Out of 3.4 million theses and dissertations searched in the 39-year period, the technology clusters report a total of 58,787 hits: the Internet with over +57,000 hits, Videotex with 450 hits, Social Media with 170 hit, Teletext with 163, BBS with 114 hits, National Projects with 107 hits, and Mainframe systems 49 hits.

The frequency of hits on the Internet and Social Media Clusters are still rising though 2007, suggesting they have not yet reached their peak, and hits on the Mainframe Cluster are few, so we should be wary of drawing any conclusions. Conversely, the Teletext, Videotex and National Project Clusters show substantial change over time and appear to have long past their peaks (Figure 3). The Videotex Cluster hits an average of 11.5 theses per year, with a high of 39 in 1985 and 1987, and 68% of the hits falling between 1981 and 1991. The Teletext Cluster hits on an average of 4.2 theses per year, with a high of 19 in 1983 and 69% of the hits falling between 1978 and 1988. The National Projects Cluster, concerning French Minitel and Canadian Telidon, while less popular, exhibits a similar profile: it hits on an average 2.7 theses per year, with a high of 15 in 1987 and 70% of the hits falling between 1982 and 1992. We observe strong positive correlations between these last three research clusters: the Pearson coefficient between videotex and teletext is 0.74, between videotex and national project is 0.73, and between teletext and national project is 0.66. In contrast, videotex has a negative correlation to the internet (-0.43) and BBS (-0.45) Clusters (which have a 0.58 correlation to each other). Table 4 contains the Pearson coefficients for each pair of keyword clusters.

Research programs as reflected in books published (Ngram)

The Ngram analysis over-counts by duplication, like the WorldCat analysis, but draws from a database orders of magnitude greater, including more publications and indexing every word in each of those publications. This Ngram analysis uses the number of books which contain cluster keywords, as opposed to the total frequency of those keywords, so as to minimize the impact of individual texts which use any of the words many times (i.e. ensuring that a single book on teletext does not massively inflate the keyword count). Table 2 lists, by year, the percent of GoogleBooks in the Ngram database containing hits for each technology cluster. Figure 4 is a stacked column graph for the period of 1970 to 1990 while Figure 5, Figure 6 and Figure 7 illustrate the explosion of internet-related keywords in the following two decades (all of which, again, dwarf the technology clusters which concern this study).

The Ngram data represented in Figure 4 shows the same rise and fall of interest in videotex and teletext which we saw in the Worldcat data, and the beginning of the Internet take-off of the mid-1990s. In corroboration of the WorldCat observations, the

Videotex, Teletext and National Projects Clusters all reached their Ngram peak in 1985 and over 50% of the related keyword hits fall between 1980 and 1990.

In contrast to the Worldcat data, however, the Mainframe and BBS clusters are more apparent. An examination of the raw counts reveals that 67,691 books hit on mainframe keywords (about equal hits each for AOL and Compuserv) and 22,046 on BBS keywords (of which 1,148 are for FidoNet). The Mainframe Cluster reached its peak in 2002 and 69% of hits fell between 1997 and 2007 (another 25% fall between 1984 and 1996). The BBS Cluster observed in the Ngram data is the most diffuse observed: the cluster reaches its peak in 1995 but only 43% of the hits fall between 1990 and 2000. The FidoNet keyword, which is likely to be more accurate than the BBS acronym, also peaks in 1995, after making its first appearance in 1985. 63% of FidoNet hits fall between 1990 and 2000 and 34% between 2001 and 2008.

Table 5 reports the correlations between keyword clusters in the Ngram data, which are similar to those observed in the WorldCat data: Pearson coefficients between videotex, teletext and National Programs are between 0.94 and 0.98, indicating an almost direct positive correlation. The Ngram data observes a much stronger correlation between the Mainframe and Internet Clusters (0.82) and a weaker correlation between the Internet and Social Media Clusters (0.46).

Interpreting the Results

Qualitative investigation of the phenomena revealed by this quantitative study reveals a number of lessons for researchers of technology and digital policy. First, as predicted by Pinch and Bijker, during the early stages of development technologies are subject to a great deal of flux in definition, meaning and naming. The early years of teletext and videotex is rife with maddening confusion about what defined each technology, what they should be called or how they should be understood. The subject of title of this paper – "What Teletex Can Teach Us" – has never been clear, because there has never been consensus upon what "teletex" means.

Second, institutions tend to develop technologies through the frame of their existing paradigm; both newspaper and television companies attempted to reproduce their business models in videotex and teletext. The institutional developers of each of the technologies expected their work product to revolutionize retailing, business services and the consumption of information by home users (BBS were looking for a better way to share messages between hobbyists). As predicted by Kuhn, these expectations persisted despite dozens of trials in which consumers failed to use services expected. Researchers can be forgiven for believing interviewees who often stated that they would be willing to use or pay for services, which they opted not to adopt when ultimately offered the choice. There are many instances, however, where researchers dismissed findings because the users sampled were considered atypical, this is usually manifested in the statement that hobbyists are not "normal". Most commonly, users who said they wanted to obtain useful information were more interested in recreation or communicating with other users, findings which (as predicted by Pinch, Bijker and Latour) only prevailed after intense conflict within institutions sponsoring the research.

Third, we should inquire why, of the research programs targeted in this study, the most widely adopted appears to have been the least studied. By 1993, the FidoNet BBS network had grown to a globe-spanning 20,000 nodes but was mentioned in only four books. BBSes had already reached a market size sufficient to support a small industry of software developers and trade publications, yet WorldCat lists only one thesis or dissertation using the FidoNet keyword between 1991 and 1995.

The last lesson is that technologies do not die: we simply stop talking about them. In June 2011, telnetBBSguide.com lists 354 is active BBSes, and Syncro.net, a BBS software publisher re-launched in 1999, lists another 133 on a separate network. FidoNet still carries new messages and Egypt recently experienced a BBS revival after the government cut off the country's connection to the Internet. The PRESTEL videotex system was purchased by the Financial Times and converted to an information services product for the financial industry, and there is reason to suspect that some videotex systems live on in the travel industry. Teletext technology persists in closed-captioning and subtitles of movies and television shows, as well as in marquee displays in airports and institutions around the world. While visiting Belgium, a Dutch football fan in his twenties gave the author a tour of the Dutch teletext service still operating in that country. This young man was born after the decline of teletext, according to the results of this study, but his enthusiasm and loyalty to the technology (which provides his morning fix of sports scores) begs for further investigation into the what role "failed" technologies may continue to play in the present day.

Tables and Figures

Figure 1: Percent of graduate studies per year per keyword cluster, (1970-1994)

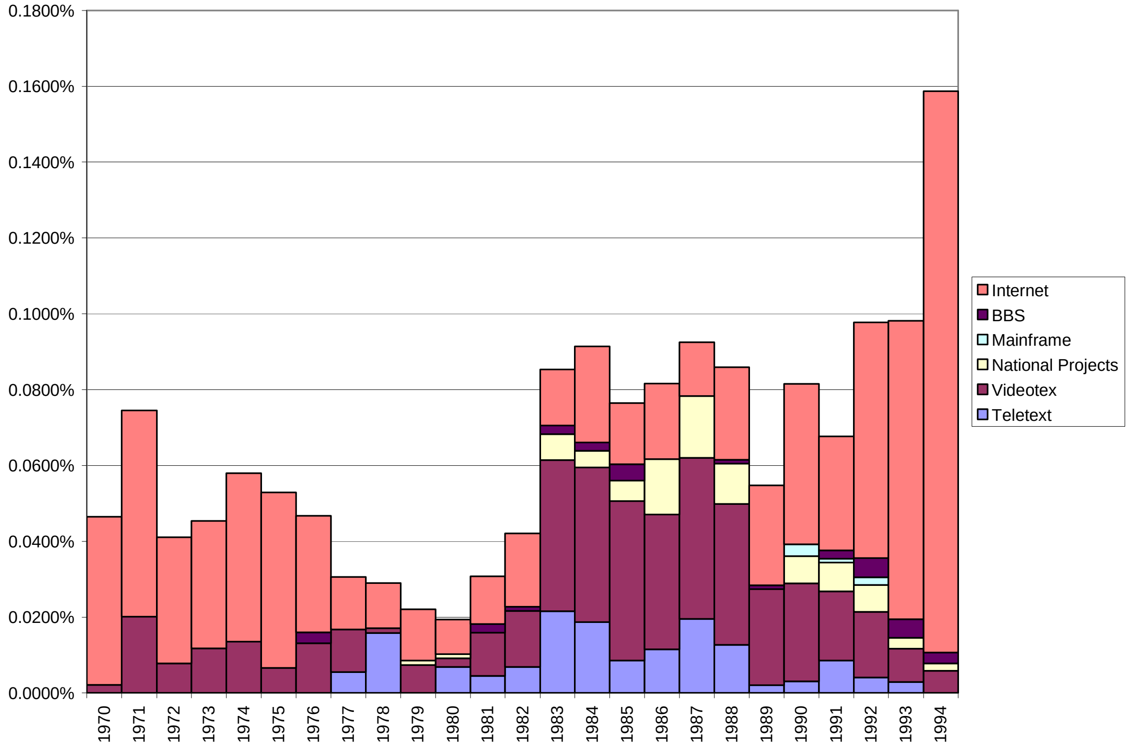

Figure 2: Percent of graduate studies per year per keyword cluster, (1970-2008)

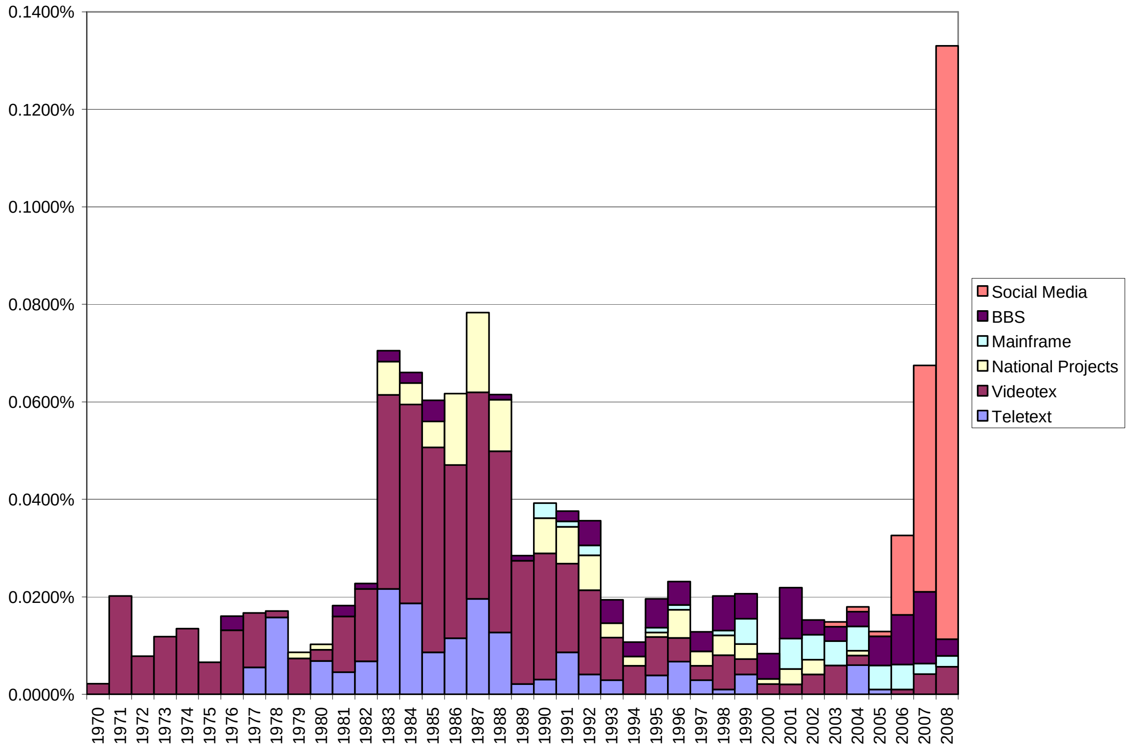

Figure 3: Percent of graduate studies using Videotex, Teletext, National Project, Mainframe or BBS keywords keywords, (1970-2008)

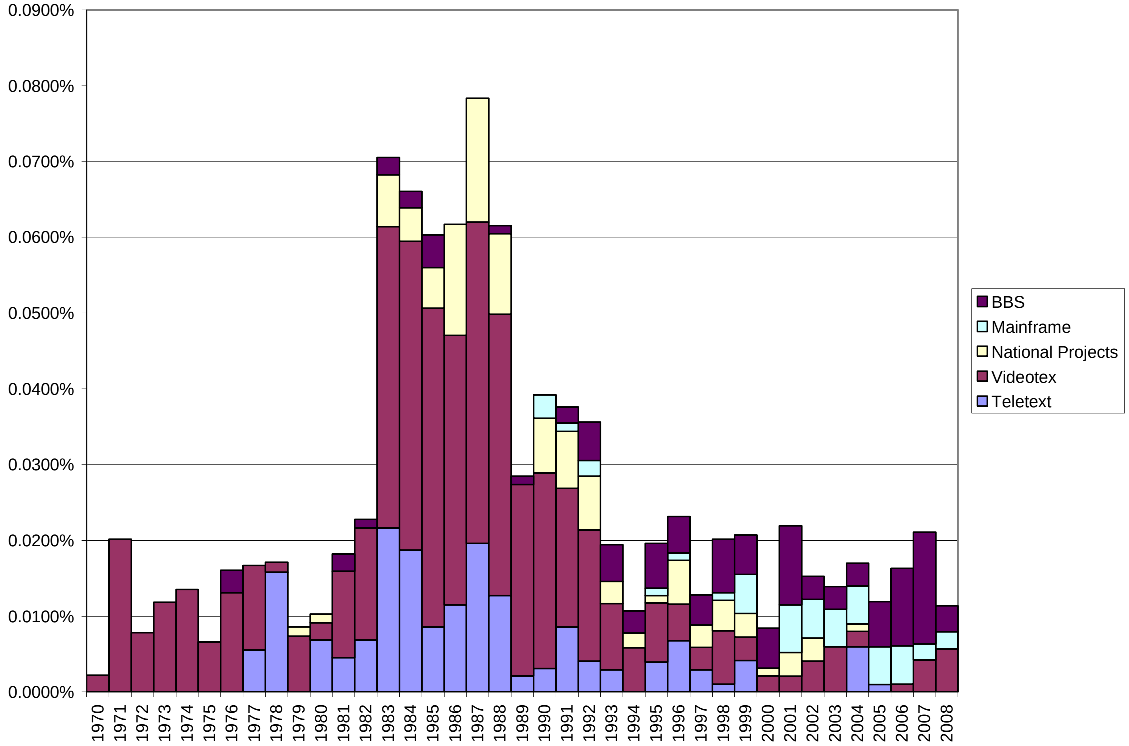

Figure 4: Percent of GoogleBooks mentioning technology clusters (1970-1990)

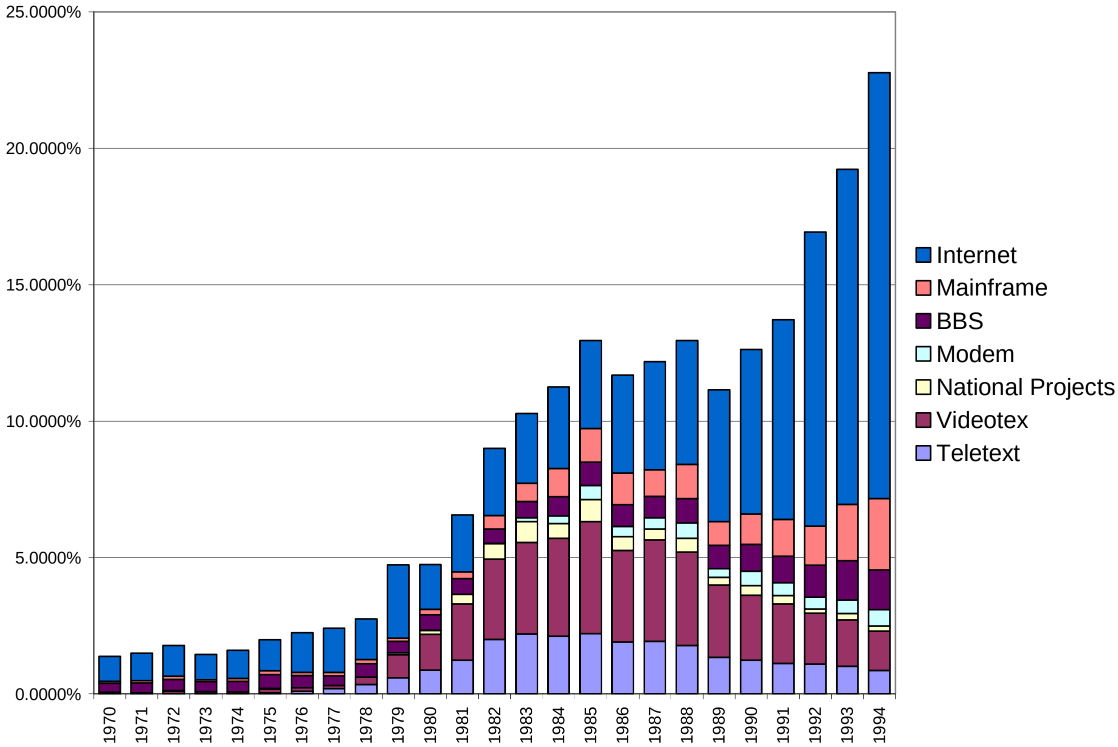

Figure 5: Percent of GoogleBooks mentioning Internet (1970-2008)

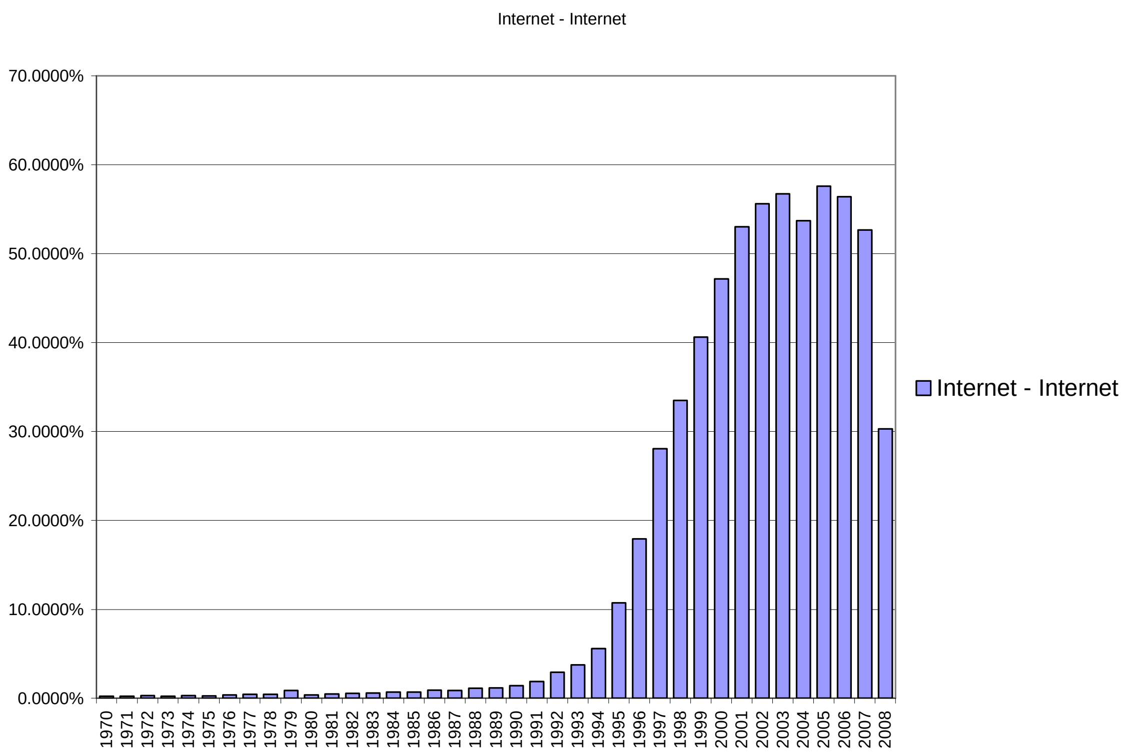

Figure 6: Percent of GoogleBooks mentioning UUNet, USENET or ARPAnet (1970-2008)

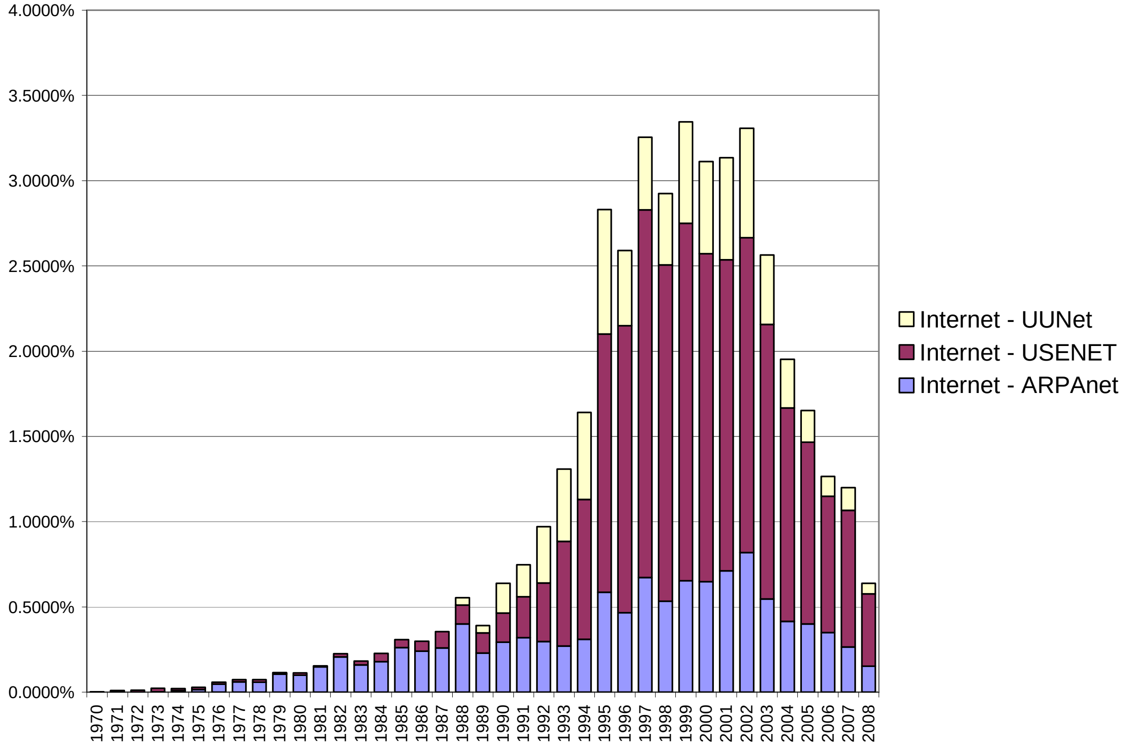

Figure 7: Percent of GoogleBooks mentioning WWW or HTTP (1970-2008)

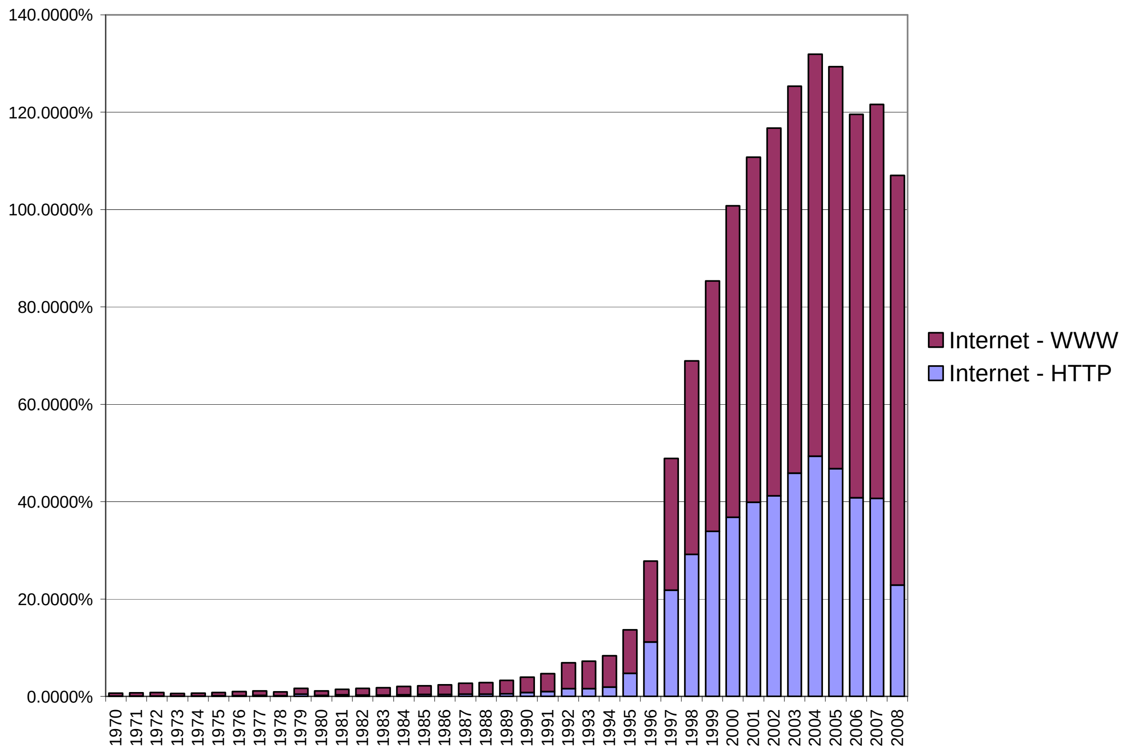

Table 1: Number of theses and dissertations per technology cluster, (1970-2008)

year	Defenses per year	Teletext	Videotex	National Projects	Mainframe	BBS	Social Media	Internet
1970	45188		1					20
1971	49633		10					27
1972	51125		4					17
1973	50694		6					17
1974	51782		7					23
1975	60544		4					28
1976	68515		9			2		21
1977	71798	4	8					10
1978	75929	12	1					9
1979	81452		6	1				11
1980	87622	6	2	1				8
1981	87771	4	10			2		11
1982	87853	6	13			1		17
1983	87904	19	35	6		2		13
1984	90787	17	37	4		2		23
1985	92813	8	39	5		4		15
1986	95627	11	34	14				19
1987	91911	18	39	15				13
1988	94276	12	35	10		1		23
1989	94889	2	24			1		25
1990	96897	3	25	7	3			41
1991	93065	8	17	7	1	2		28
1992	98229	4	17	7	2	5		61
1993	102889	3	9	3		5		81
1994	102700		6	2		3		152
1995	102145	4	8	1	1	6		311
1996	103660	7	5	6	1	5		691
1997	101610	3	3	3		4		1,378
1998	99141	1	7	4	1	7		2,280
1999	96688	4	3	3	5	5		2,676
2000	95115		2	1		5		3,938
2001	95813		2	3	6	10		4,187
2002	98291		4	3	5	3		4,441
2003	100697		6		5	3	1	4,841
2004	100162	6	2	1	5	3	1	5,611
2005	100604	1			5	6	1	5,698
2006	98043		1		5	10	16	6,520
2007	94820		4		2	14	44	7,326
2008	87962	-	5	-	2	3	107	7,123

Source: WorldCat.org database, May 2011.

Table 2: Percent of GoogleBooks per technology cluster keyword, (1970-2008)

year	Teletext	Videotex	National Projects	Mainframe	BBS	Social Media	Internet	year
1970		0.0607%		0.0767%	0.3293%		0.9047%	1970
1971		0.0439%		0.0777%	0.3582%		1.0170%	1971
1972		0.1072%		0.1137%	0.4159%		1.1308%	1972
1973	0.0136%	0.0781%		0.0747%	0.3634%		0.9170%	1973
1974	0.0104%	0.0659%		0.1109%	0.3848%		1.0226%	1974
1975	0.0462%	0.1385%	0.0165%	0.1286%	0.5144%		1.1375%	1975
1976	0.1044%	0.1175%		0.1109%	0.4568%	0.0098%	1.4521%	1976
1977	0.1898%	0.1244%		0.1244%	0.3502%	0.0098%	1.6265%	1977
1978	0.3435%	0.2748%		0.1505%	0.4906%		1.4981%	1978
1979	0.5884%	0.8506%	0.0672%	0.1151%	0.4253%	0.0128%	2.6798%	1979
1980	0.8784%	1.3127%	0.1371%	0.2024%	0.5616%	0.0163%	1.6457%	1980
1981	1.2350%	2.0676%	0.3538%	0.2498%	0.5689%		2.0780%	1981
1982	2.0071%	2.9383%	0.5561%	0.4968%	0.5232%		2.4645%	1982
1983	2.1964%	3.3540%	0.7621%	0.6689%	0.5949%		2.5565%	1983
1984	2.1158%	3.5912%	0.5403%	1.0291%	0.7043%	0.0091%	2.9871%	1984
1985	2.2023%	4.1165%	0.7981%	1.2332%	0.8551%		3.2254%	1985
1986	1.9044%	3.3530%	0.5051%	1.1612%	0.7954%		3.5881%	1986
1987	1.9341%	3.7146%	0.3982%	0.9727%	0.7907%	0.0057%	3.9592%	1987
1988	1.7714%	3.4329%	0.5081%	1.2468%	0.9035%	0.0055%	4.5397%	1988
1989	1.3467%	2.6512%	0.2799%	0.8714%	0.8635%	0.0053%	4.8323%	1989
1990	1.2388%	2.3804%	0.3498%	1.1076%	0.9983%	0.0049%	6.0335%	1990
1991	1.1207%	2.1805%	0.3050%	1.3495%	0.9708%	0.0051%	7.3217%	1991
1992	1.0883%	1.8748%	0.1558%	1.4292%	1.1784%	0.0049%	10.7738%	1992
1993	1.0131%	1.6950%	0.2372%	2.0632%	1.4504%	0.0025%	12.2829%	1993
1994	0.8649%	1.4341%	0.1972%	2.6217%	1.4520%	0.0045%	15.5959%	1994
1995	1.0034%	1.5294%	0.1746%	4.2169%	2.0024%	0.0044%	27.2488%	1995
1996	0.7033%	1.2348%	0.1819%	5.3493%	1.8511%	0.0081%	48.3237%	1996
1997	1.0118%	1.4205%	0.1100%	6.8823%	1.7427%	0.0098%	80.2216%	1997
1998	0.8209%	0.9818%	0.0804%	6.5800%	1.5321%		105.3697%	1998
1999	0.8208%	1.1211%	0.1388%	7.3494%	1.5185%	0.0052%	129.3003%	1999
2000	0.7132%	1.0860%	0.0954%	7.1350%	1.4132%	0.0249%	151.0691%	2000
2001	0.8140%	0.9907%	0.0724%	7.7222%	1.4192%	0.0696%	166.9074%	2001
2002	0.6051%	0.9506%	0.0750%	7.7334%	1.3616%	0.1155%	175.6504%	2002
2003	0.6310%	0.8165%	0.0701%	6.8601%	1.2722%	0.2155%	184.5911%	2003
2004	0.5110%	0.7830%	0.0655%	5.2775%	1.1040%	0.4796%	187.5976%	2004
2005	0.5461%	0.7763%	0.0414%	4.5267%	1.2415%	0.8890%	188.5638%	2005
2006	0.3838%	0.7501%	0.0441%	3.8302%	1.1650%	1.9215%	177.1972%	2006
2007	0.3590%	0.5541%	0.0258%	3.5839%	1.1447%	4.6681%	175.4820%	2007
2008	0.1091%	0.2798%	0.0134%	1.8591%	0.8141%	4.6863%	137.9279%	2008

Source: GoogleBooks Ngram database (googlebooks-eng-all-20090715)

Table 3: Composition of technology keyword clusters in Ngram database, (1970-2008)

	Teletext Cluster			Videotex Cluster			
year	CEEFAX	Teletext	Videotext	PRESTEL	Teletex	Videotex	Viewdata
1970				0.0607%			
1971				0.0439%			
1972				0.1040%			0.0032%
1973		0.0136%		0.0781%			
1974	0.0104%			0.0659%			
1975	0.0066%	0.0396%		0.0758%	0.0033%	0.0297%	0.0297%
1976	0.0522%	0.0457%	0.0065%	0.0816%		0.0033%	0.0326%
1977	0.0687%	0.1113%	0.0098%	0.0720%		0.0098%	0.0425%
1978	0.1079%	0.2355%		0.1014%	0.0033%	0.0098%	0.1603%
1979	0.1855%	0.3773%	0.0256%	0.3646%	0.0192%	0.0863%	0.3805%
1980	0.1861%	0.5747%	0.1176%	0.4996%	0.0424%	0.2286%	0.5420%
1981	0.3053%	0.7285%	0.2012%	0.8048%	0.1006%	0.5030%	0.6591%
1982	0.3290%	1.2306%	0.4475%	0.9739%	0.2205%	0.7798%	0.9641%
1983	0.2412%	1.3120%	0.6431%	0.9165%	0.2766%	1.2413%	0.9197%
1984	0.2823%	1.1414%	0.6921%	0.9623%	0.3370%	1.4753%	0.8166%
1985	0.2550%	1.2242%	0.7231%	0.9901%	0.4231%	1.8842%	0.8191%
1986	0.1829%	1.0393%	0.6822%	0.6764%	0.3658%	1.6866%	0.6241%
1987	0.1763%	1.1206%	0.6371%	0.7793%	0.4352%	1.7975%	0.7025%
1988	0.1373%	1.0052%	0.6289%	0.6756%	0.4998%	1.6313%	0.6262%
1989	0.1294%	0.7473%	0.4700%	0.5836%	0.3354%	1.2490%	0.4832%
1990	0.1360%	0.6291%	0.4736%	0.5562%	0.3303%	1.1950%	0.2988%
1991	0.0915%	0.5743%	0.4549%	0.5057%	0.2872%	1.1055%	0.2821%
1992	0.0609%	0.6209%	0.4066%	0.4188%	0.2849%	0.9690%	0.2021%
1993	0.0494%	0.4794%	0.4843%	0.4003%	0.2520%	0.8846%	0.1581%
1994	0.0717%	0.3787%	0.4145%	0.4526%	0.1613%	0.7081%	0.1120%
1995	0.0486%	0.4663%	0.4884%	0.4973%	0.1459%	0.7691%	0.1171%
1996	0.0667%	0.3557%	0.2809%	0.4547%	0.1273%	0.5396%	0.1132%
1997	0.0786%	0.6012%	0.3320%	0.5855%	0.1100%	0.5953%	0.1297%
1998	0.0768%	0.4296%	0.3145%	0.5211%	0.0622%	0.3419%	0.0567%
1999	0.0937%	0.4408%	0.2863%	0.5692%	0.0885%	0.3662%	0.0972%
2000	0.0925%	0.4036%	0.2172%	0.5973%	0.0587%	0.3229%	0.1071%
2001	0.1030%	0.4800%	0.2310%	0.5941%	0.0570%	0.2727%	0.0668%
2002	0.0679%	0.3788%	0.1584%	0.5694%	0.0393%	0.2895%	0.0524%
2003	0.0660%	0.4124%	0.1526%	0.5629%	0.0278%	0.1639%	0.0619%
2004	0.0701%	0.2961%	0.1448%	0.5681%	0.0378%	0.1411%	0.0360%
2005	0.0597%	0.3487%	0.1377%	0.5663%	0.0289%	0.1397%	0.0414%
2006	0.0376%	0.2295%	0.1166%	0.5637%	0.0239%	0.1267%	0.0358%
2007	0.0428%	0.2298%	0.0864%	0.4401%	0.0160%	0.0811%	0.0169%
2008	0.0147%	0.0649%	0.0295%	0.2269%	0.0080%	0.0328%	0.0121%

	National Projects Cluster		Mainframe Cluster		BBS Cluster	
year	Minitel	Telidon	AOL	Compuserv	BBS	FidoNet
1970			0.0767%		0.3293%	
1971			0.0777%		0.3582%	
1972			0.0715%	0.0422%	0.4159%	
1973			0.0679%	0.0068%	0.3634%	
1974			0.1109%		0.3848%	
1975	0.0033%	0.0132%	0.1286%		0.5144%	
1976			0.1109%		0.4568%	
1977			0.1047%	0.0196%	0.3502%	
1978			0.1439%	0.0065%	0.4906%	
1979	0.0032%	0.0640%	0.1087%	0.0064%	0.4253%	
1980		0.1371%	0.1469%	0.0555%	0.5616%	
1981	0.0069%	0.3469%	0.1457%	0.1041%	0.5689%	
1982	0.0033%	0.5528%	0.1415%	0.3554%	0.5232%	
1983	0.0096%	0.7525%	0.1351%	0.5338%	0.5949%	
1984		0.5403%	0.2004%	0.8287%	0.7043%	
1985	0.0600%	0.7381%	0.2040%	1.0291%	0.8521%	0.0030%
1986	0.0523%	0.4529%	0.1771%	0.9841%	0.7954%	
1987	0.1081%	0.2901%	0.1593%	0.8134%	0.7736%	0.0171%
1988	0.1730%	0.3351%	0.1922%	1.0546%	0.8733%	0.0302%
1989	0.1426%	0.1373%	0.1611%	0.7103%	0.8503%	0.0132%
1990	0.1700%	0.1797%	0.1506%	0.9570%	0.9570%	0.0413%
1991	0.1779%	0.1271%	0.2186%	1.1309%	0.9378%	0.0330%
1992	0.0828%	0.0730%	0.1826%	1.2466%	1.0835%	0.0950%
1993	0.1730%	0.0642%	0.2446%	1.8186%	1.2453%	0.2051%
1994	0.1389%	0.0583%	0.3384%	2.2834%	1.2750%	0.1770%
1995	0.1525%	0.0221%	0.6255%	3.5915%	1.7129%	0.2895%
1996	0.1213%	0.0606%	1.0104%	4.3389%	1.6693%	0.1819%
1997	0.0982%	0.0118%	1.6425%	5.2398%	1.5717%	0.1709%
1998	0.0475%	0.0329%	2.0148%	4.5652%	1.4279%	0.1042%
1999	0.0868%	0.0521%	2.5389%	4.8105%	1.4039%	0.1145%
2000	0.0631%	0.0323%	3.0847%	4.0503%	1.3222%	0.0910%
2001	0.0417%	0.0306%	4.3105%	3.4117%	1.3316%	0.0877%
2002	0.0417%	0.0334%	4.9733%	2.7600%	1.2972%	0.0643%
2003	0.0433%	0.0268%	4.7548%	2.1052%	1.2114%	0.0608%
2004	0.0378%	0.0277%	3.7713%	1.5061%	1.0671%	0.0369%
2005	0.0318%	0.0096%	3.2756%	1.2511%	1.1558%	0.0857%
2006	0.0312%	0.0129%	2.8139%	1.0163%	1.1311%	0.0340%
2007	0.0151%	0.0107%	2.7706%	0.8133%	1.1180%	0.0267%
2008	0.0067%	0.0067%	1.4306%	0.4285%	0.7947%	0.0194%

Source: GoogleBooks Ngram database (googlebooks-eng-all-20090715)

Table 4: Table of Correlations for WorldCat data

Table of Correlations for WorldCat data

	Teletext	Videotex	National Projects	Mainframe	BBS	Social Media	Internet
Teletext		0.64	0.59	-0.18	-0.49	-0.63	-0.44
Videotex	0.64		0.75	-0.44	-0.45	0.47	-0.45
National Projects	0.59	0.75		-0.31	-0.34	-1.00	-0.47
Mainframe	-0.18	-0.44	-0.31		0.07	-0.83	0.48
BBS	-0.49	-0.45	-0.34	0.07		0.02	0.58
Social Media	-0.63	0.47	-1.00	-0.83	0.02		0.85
Internet	-0.44	-0.45	-0.47	0.48	0.58	0.85	

Table 5: Table of Correlations for NGram data

	Teletext	Videotex	National Projects	Mainframe	BBS	Social Media	Internet
Teletext		0.98	0.95	-0.16	0.01	-0.44	-0.36
Videotex	0.98		0.94	-0.12	0.10	-0.37	-0.30
National Projects	0.95	0.94		-0.53	-0.42	-0.43	-0.63
Mainframe	-0.16	-0.12	-0.53		0.83	-0.01	0.82
BBS	0.01	0.10	-0.42	0.83		-0.08	0.54
Social Media	-0.44	-0.37	-0.43	-0.01	-0.08		0.46
Internet	-0.36	-0.30	-0.63	0.82	0.54	0.46	