JAY HAUBEN

Libraries of the Future 1945 – 1965

Questions from Vannevar Bush, John Kemeny and JCR Licklider[1]

Throughout history thinkers and scholars have lamented that there is not enough time to read everything of value. The real problem is not the volume of valuable scholarship and recorded thought and reasoning. The historic problem for scientists and scholars has been selecting and gathering the relevant material and processing it in their own brains to yield new knowledge. The goal is to contribute new insights to the body of knowledge, to enhance what we have to draw on and what gets passed on from generation to generation in addition to biologically inherited genetic information.

A grand vision emerged in the US after the Second World War. New human-machine knowledge management systems would be developed to help researchers consult more of the corpus of all recorded knowledge. Such systems would increase the usefulness of the corpus and accelerate the making of new contributions to it.

1. *Vannevar Bush and the Memex*

Vannevar Bush (1890 – 1974), an American inventor, engineer and science administrator is popularly considered to have initiated this vision in July 1945 with his article "As We May Think"[2]. In the 1920s and 1930s, Bush had designed and built the first large scale analog computers. These were used to solve differential equations, being an advanced use of machines to do mental work. During the Second World War, Bush had directed the US Office of Science Research and Development which managed and coordinated the war related activities of some

1 The following is a slightly revised version of a talk presented on March 27, 2004 at the "Wissensmanagement in der Wissenschaft" conference in the Institute for Library Science of Humboldt University in Berlin, co-sponsored by the Institute for Library Science and the Society for Science Research.

2 Bush, V., As We May Think. – In: The Atlantic Monthly (Boston). 176 (1945), 1. 101 – 108. Online at http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm. Reprinted including illustrations from Life Magazine in Nyce, J.M. and Kahn, P., (ed), From Memex to Hypertext: Vannevar Bush and the Mind's Machine. Boston: Academic Press 1991. p. 85 – 110.

6000 US scientists. As the end of the war was coming into sight, Bush saw two problems emerging:

1) how to make the huge volume of war time reports and research findings public and accessible and
2) what new challenge to set for the scientists who would be finishing their war related work.

His article "As We May Think" proposed one solution for both problems. Bush proposed the development of mechanical systems to manage and process the growing body of scientific, technical and scholarly information and knowledge.

Bush had great faith in the lasting benefit to human society of scientific and technical development. He welcomed the growing mountain of research. The record must continue to be extended, it must be stored and above all it must be consulted and built upon. To Bush the difficulty was that "publication has been extended far beyond our present ability to make real use of the record." He worried with so much research and the necessary specialization that "significant attainments become lost in the mass of the inconsequential."

But there were signs of hope. Bush was at heart a great inventor. He offered as a solution a desk-like device he called "memex", (perhaps for *mem*ory *ex*tension). It would be a mechanized file and personal library system. Using improved microfilm, it would have the capacity to store all the books, documents, pictures, correspondence, notes, etc. that a scholar or scientist might need. The microfilm texts would be created by the scholar or received in the mail from colleagues or purchased from publishers or other information providers. The cost would be minimal because the microfilm and mail would be inexpensive. Since the memex would have the capacity to dry photograph whatever the user wrote or placed on its transparent writing surface there was practically no limit to what the scholar could have available. There would be no problem storing even a million books on microfilm in a small space inside the memex. A mechanized rapid selector based on a single frame as an item would allow the call up of any frames or items desired in a very short time. The scholar's work would be facilitated by his or her own personal complete and frequently updated memex library.

But what good is all this personal accumulation of the record? The real heart of the matter for the scholar is to find in the corpus what is relevant and intellectually stimulating. The problem Bush saw that needed to be solved was the method of selection. So far, indexing and cataloguing were done alphabetically or numerically and searching or selecting was by tracing down from subclass to subclass. For example in consulting a dictionary or an index, the first letter is found, then the second, and so on. Such a method Bush wrote was artificial. The human brain does not work that way.

The essence of the memex would be to store, organize and retrieve in a way analogous to the working of the brain. How does the human brain work? It operates, according to Bush's understanding, by *association*. Describing the working of the human brain, Bush observed, "With one item in its grasp, [the brain] snaps instantly to the next that is suggested by the association of thoughts." This is "in accordance with some intricate web of trails carried by the cells of the brain."[3] Recall is sometimes vague and trails not frequently followed are prone to fade with time. Yet the brain is awe-inspiring with its speed of action, intricacy of details and recall of mental pictures.

How could the memex act like the brain? Every time the scholar or scientist puts the microfilm of a book or document into the memex he or she assigns to it a code in the codebook section of the memex. That is the same as before. But, in imitation of the brain, every time the scholar consults a document or item in the memex, the scholar has a mechanism to associate it with other items which come to mind. From then on, the associated items will be able to select each other automatically. The memex puts codes in the margin of the microfilm to insure this action. As the user consults an item in the memex or does his or her scholarly work, trails of association are thus created and recorded for later use. The contents of the memex are in this way organized and coded for retrieval or further research. Every item consulted is associated with other items that are intellectually connected with it. Selection by association replaces indexing. The scholar can annotate the trails, draw conclusions from them and when satisfied that something worthwhile has been discovered have the memex make copies of the trail and the documents associated with it. The memex makes the copies photographically on microfilm, in the process a new document is made of the associated frames. The scholar can send the associative trail to his colleagues for insertion of it into their own memexes to be combined with their own trails or the scholar can send it to a publisher for publication.

Bush expected in this way to increase the accessibility and utility of the store of knowledge customized by each user and to facilitate collaboration and dissemination of new knowledge. He also expected, in time, ways would be found so that each memex would learn from the usage of each scholar how to increase the usefulness of its operation. Eventually advanced memexes could be instructed to search for new trails that would be useful to the scholar but which he or she had not yet discovered. In essence, Bush's associative trails were a new knowledge structure and a memex memory coded with associative indexing a new memory structure. Bush expected wholly new forms of encyclopedias would be made,

3 Ibid., Nyce, J. / Kahn, P., p. 101.

with a mesh of associative trails running through them. A new profession of trail-blazers would appear for those who took pleasure in finding useful trails through the enormous mass of the common record. By the easy exchange of microfilmed trails, Bush was hopeful scholarly collaboration and co-work would be facilitated and become common.

Bush expected, having modeled the memex on the working of the brain, the memex would facilitate and accelerate scholarly and scientific work. The users of the memex also might improve their own mental processes via its use. The benefit from use of the memex would be achieved without unduly adding to the cost of storage or dissemination because the memex would cause scholarly and scientific publishing to change to microfilm as well. Bush was hopeful in 1945 that the improved knowledge management introduced by memex might yet allow everyone to "encompass the great record and to grow in the wisdom" of human experience.[4]

There is little evidence a memex was ever built. Digitalization replaced microfilm and all-purpose electronic computers became available so that microfilm and photographic methods were no longer considered as the basis for a scholarly workstation. But the idea of associative trails or associative indexing is often sited as the inspiration for hypermedia knowledge structures that have proliferated since the early 1990s. Whether the memex would have ever lived up to Bush's expectations, Bush used it to raise important questions for knowledge management for the sciences: How can the whole corpus of knowledge in a scientist's field be made available to him or her and be kept current? How should it be organized? What method of search and retrieval? And how can knowledge be shared and collaboratively generated? Bush also pointed in the intriguing direction. Look to the master of knowledge management, the human brain for help with knowledge management.

2. *John Kemeny and the National Research Library*

The questions and prospects raised by Vannevar Bush in 1945 especially about automation and information handling were taken seriously in the community of scholars around the Massachusetts Institute of Technology, in particular in the cybernetic circles.[5] In 1961, MIT celebrated its 100th anniversary. Among other events, a series of eight lectures were planned addressing the topic "Management

4 Ibid., Nyce, J. / Kahn, P., p. 107.

5 Hauben, R. Cybernetics, Time-Sharing, Human-Computer Symbiosis and Online Communities. – In: Hauben, M. / Hauben, R., Netizens: On the History and Impact of Usenet and the Internet. Los Alomitos, Ca: IEEE Computer Society Press 1997. p. 76 – 95.

and the Computer of the Future”. Many from the cybernetics community attended the lectures and the discussions went far beyond the question of management. The final gathering of talks appeared in book form under the broader title Computers and the World of the Future.[6]

John G. Kemeny (1926 – 1992), the Hungarian born mathematician and co-creator of the BASIC computer language and the Dartmouth Time Sharing operating system, gave one of the MIT lectures. His presentation on a “Library for 2000 AD”[7] was followed by a lively panel and audience discussion. Kemeny’s presentation echoed the concern since the 1930s that US research libraries were facing the problems of rapid growth leading to increasing difficulty to manage and use such libraries. He gave as an example of a difficulty a common practice. If a book is misplaced on the shelves in a major library and cannot be located after a short search, it is less costly to replace the book than to continue the search. He argued that keeping up to date with research and publication even in a scientist’s own subfield was growing ever more difficult. Relevant previous or current work is easy to miss. Kemeny drew the conclusion that the research library had to be radically reorganized.

Surely, automation could play a big role in the reorganization, but Kemeny warned not to use a machine where a human can perform the same task better or more efficiently. Also, because books are “most inconvenient for machine processing,” Kemeny, like Bush foresaw use of another medium, for example magnetic tape or photographic microfilm. To have the whole corpus of scholarly and scientific books available to the whole science and research communities, Kemeny proposed a single library centralized or maybe diffused, where all research material on tape or film would be house and made accessible over telephone lines. He called it the “National Research Library”, the NRL. He would not abandon book libraries; only reduce them each to no more than a few hundred thousand reference, leisure reading and core research books in all fields. The space freed up he would use for study rooms and reading rooms equipped with terminal, tape readers and printout devices.

Kemeny proposed dividing all research material into subjects, and all subjects into branches and subbranches. Each user would be guided to do research in one of the subbranches. The whole body of recorded material on tape or film for each branch or subbranch would be assembled into one room of the National Research Library. Each room would be part of a comprehensive human+selection-

6 Greenberger, M., (ed.), Computers and the World of the Future. Cambridge, MA.: The MIT Press1962.

7 Ibid., p. 134 – 178.

machine+computer system. The system would be based on chapter or article length items. Expert human reviewers would assign each item to what the reviewer judged its appropriate subbranch and therefore room. Expert abstractors at the NRL or maybe the author or book reviewer would write a detailed text abstract for each item including in the abstract all bibliographic information and all citations in the item. The abstracts would be on the tape or the film along with the items perhaps in code in the margins to facilitate searching. The search would be in the database of coded abstracts. When a search was finished, the complete item or items matching the search criteria would be retrieved, converted to a form transmittable over phone lines and sent to the user who had requested the search.

The scholar or scientist would sit at a terminal in the home institution or office. He or she would use a telephone system to dial up the branch or subbranch to search in and would connect to a computer system programmed to help delimit the search. The “conversation” on the terminal screen between scholar and computer would be a give and take. Questions provided to the computer program by subject experts would guide the scholar in narrowing the search to a manageable level. At some point the computer program would judge the range of the search would yield only a few thousand abstracts. It would signal the scholar to wait while it tries to cluster the matching abstracts by some statistically discovered shared features. The computer would then display the features that it discovered for the different clusters. The scholar would make the final judgment of relevancy. The 20 to 200 abstracts so chosen would then be copied with their items onto a tape or film. The collection might be scanned by a video camera and transmitted by the telephone to the scholar’s terminal for printout and study. One of the commentators at the end of the lecture pointed out that this would require rewriting copyright laws.

Kemeny would also have the library machine system keep track of its own operation and have mechanisms for adapting its procedures as it learned from its use.

Kemeny concluded that to be able to sit in his office and rapidly get copies of all the materials he needed for his research would be attractive to him. If all scholars had such access to all the resources of the NRL whenever they needed them, there would be a great impact on the productivity of scholarship. But also, would not the nature of publishing change? Publishing could be accelerated by having all articles and book manuscripts submitted directly to the National Research Library with its staff of experts for faster review and appearance in the corpus. Each scholar could also subscribe to all new material in his or her field which would be gathered by the NLR staff or by the library machine itself and delivered once a month by a simple phone call. This human-computer system Kemeny offered as the Library for 2000 AD.

The lecture by Kemeny was followed by a panel and audience discussion. The panelists and audience questioned most of Kemeny's presentation. Such a prominent role as Kemeny gave to human experts was challenged. Experts would not necessarily agree among themselves. Who would choose the experts? How could bias by minimized? Gathering the whole corpus made sense. The method of selection was the problem. Wouldn't it be better to return as a search result all citations in a relevant item and all citations of that item since its appearance?[8] Such trails are how many scholars come to their insights. Also, dividing information and human thought into branches and subbranches losses all the knowledge that comes from cross-disciplinary thinking. Classification into subjects may be counter productive. Is not cross-referencing, not dissection the essence of a library?

Other questions were raised in the discussion. What is the essence of new knowledge? "New ideas which are really the object of information retrieval," said one commentator, "result from the inverse of a tree, namely the combination of ideas." Also, abstracts may not prove to be a good basis for a search. They are filters that remove just the subtle details that a scholar needs for new insights. Maybe what we should seek is a way to represent knowledge so that the search result would be the construction of knowledge like a theorem-proving machine does from axioms. Where in this scheme is the role of suggestions made by colleagues and librarians and the provision for collaborative work? The discussion ended with a reminder that fancy cumbersome machines for retrieving and viewing information might not be as successful as books regardless of all the faults of the book as information container. No consensus was reached.[9]

Kemeny like Bush had sought a scheme to make available to all scholars the whole corpus of recorded thought. He elsewhere suggested if home terminals were available every home with one would be a mini-university.[10] To the human-machine question and to the question of semantic searching Kemeny gave highest priority to the human expert. He foresaw that humans and computers would have a give-and-take interactivity to define a search and to judge relevancy of the search results. But the responses after his lecture suggested that Kemeny had not added much to the solution of the retrieval problem. He had however provided the basis for a discussion of how that problem might be solved. And beyond Bush he foresaw that telecommunications networks would play an important role in the library

8 This suggestion echoes Garfield, E., Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. – In: Science (Washington, D.C.). 122(1955)3159. p. 108 – 111.

9 Ibid, Greenberger, M., (ed.),Computers and the World of the Future. Cambridge, MA.: The MIT Press1962.p. 162 – 177.

10 Kemeny, J.G., Man and the Computer. New York: Charles Scribner's Sons 1972. p. 84.

of the future. Where Bush had seen the need to gather the whole library into the desk of each scholar or scientist, Kemeny saw the value of making the whole corpus of knowledge available from a shared source, the National Research Library.

3. *Licklider and the Procognitive System*

At about the same time as the presentation by Kemeny at MIT, JCR Licklider was recruited to lead a project to inquire into the application of newer technologies to information handling. JCR Licklider (1915 – 1990) was a physio-psychologist by training. For his PhD in 1942 he had mapped for the first time the different sites in the brains of cats where stimuli from sounds of different frequencies are received. Licklider had also been part of the Wiener cybernetics circles around MIT and had been one of the first people to sit at the console of a mini computer, the PDP1 and operate it in an interactive mode. The Council on Library Resources which recruited Licklider had been founded and funded by the US Ford Foundation in 1956 to address the question how could technology help libraries gather, index, organize, store and make accessible the growing body of recorded information despite the intellectual explosion of the Twentieth Century.

Licklider's project was undertaken at Bolt Beranek and Newman (BBN), the science and technology firm. BBN later became famous for its role in designing and implementing the subnetwork of the US government's ARPANET experiment. Licklider gathered at BBN a small team of engineers and psychologists supplemented by some of his colleagues at MIT.[11] For two years, 1961 – 1963, they explored "concepts and problems of libraries of the future". Licklider wrote a summary report of the project which appeared as the book, Libraries of the Future, in 1965.[12]

Licklider and his team foresaw that the whole corpus of recorded thought, at least in the sciences, law, medicine, technology and the records of business and government could sooner or later be gathered into a single central or distributed computer processable memory system. The BBN study he directed was undertaken to answer the question how might this whole corpus of recorded solid thought be organized and made accessible so that it would be attractive to use and a powerful lever for human progress.

11 At BBN: Fisher S. Black, Richard H. Bolt, Lewis C. Clapp, Jerome I. Elkind, Mario Grignetti, Thomas M. Marill, John W. Senders, and John A. Swets. From MIT: John McCarthy; Marvin Minsky, Bert Bloom, Daniel G. Bodrow, Richard Y. Kain, David Park, and Bert Raphael.

12 Licklider, JCR, Libraries of the Future. Cambridge, MA.: The MIT Press 1965.

Licklider began his report with an estimate of the size that the corpus of scientific and scholarly knowledge would be in the year 2000. His estimate was of the order of 10^{14} bytes. There seemed in 1965, and there seems today, no technical obstacle to gathering a memory system of this size. In terms of today's hardware, 500 hard drives each capable of storing 200 GB of data would suffice to hold the whole body of recorded solid thought. And there seems no obstacle yet to being able to process in a reasonable time this corpus in any way chosen.

Licklider projected that if it were found possible to process the body of recorded thought so as to have more direct access to its knowledge content, then there would be the basis of a new library system. Such a system would consist of terminals and computers and networks that would make the body of human knowledge available for all possible human needs and automatic feedback machine control purposes. Licklider chose the name „procognitive" for the system he was envisioning. Procognitive because it would be a system for the advancement and application of knowledge. Rather than being based on collections of documents and tags and retrieval methods, the procognitive system would be based on the three elements, the corpus of knowledge, the question, and the answer. There would be no transportation of matter, no books, just (1) processing of information into knowledge and (2) processing of questions into answers, all done digitally. From this point of view, authors and scientists are not seen as contributing documents to science or the procognitive system. They contribute information or their thoughts which get processed for their knowledge content, augmenting the already existing corpus of knowledge.

How could information be processed into knowledge? How should the corpus of knowledge be organized? Like Bush, Licklider looked to the brain. He recognized that the human brain is a complex arrangement of neuronal elements and processes. These elements and processes "accept diverse stimuli, including spoken and printed sentences and somehow process and store them in ways that support the drawing of inferences and the answering of questions."[13] The human brain (1) processes stimuli at the time of input and (2) stores, not the stimuli but a representation of them. The inferences and answers arrived at by the brain are not mere restatements of past inputs drawn from memory but are tailored to be appropriate to the actual or current need. Licklider also believed, in part, that humans think by "manipulating, modifying, and combining 'schemata,'"[14] or schemes and models of how things work or relate to each other. New knowledge he believed is achieved by adapting one or more old schemata to fit new situations.

13 Ibid., Licklider, JCR., p. 24 – 25.

14 Ibid., Licklider, JCR., p. 3.

Could the body of thought be processed into a new body of knowledge schemata or other knowledge structures? If so, then queries of it could be answered with knowledge structures as answers rather than with already existing documents or parts of documents.

Licklider saw as the aim of the procognitive system to enable a researcher or scholar, or eventually anyone, to present to the system a search prescription or query or question in more or less natural language and get in return "suggestions, answers to questions, and made-to-order summaries" gathered from the knowledge structures in the corpus of knowledge. The outputs would not be reproductions or mere translations of previous inputs. Licklider expected the outputs to be "of the kind that a good human [research] assistant might prepare if he [or she] had a larger and more accurate memory and could process information faster."[15]

Licklider's BBN project considered or experimented with relational nets, syntactic analyses, the possibility of semantic nets, knowledge "representation languages" and other structures. Based on his sense of how the brain worked, Licklider in the early 1960s considered finding a representation language the most promising way forward. Research was needed to discover the form of the language representation that would be the foundation of a question answering system. Then computer programs and human-computer systems could be worked out that could process the whole corpus of thought and information into the representation or representations that would best capture the knowledge content of the corpus. Licklider expected such a representation language would be more rule bound than natural language, less ambiguous and would require a larger memory than the natural language text and images based corpus requires.

After the whole corpus of text and images was processed into the chosen knowledge representation form, any new contribution would be similarly processed before it would be added to the processed corpus. This processing even with the most advanced programming would require human-computer interaction. The processing would have to be organized, controlled, monitored and corrected by workers in a new profession, the procognitive "system specialists." For example, the system would issue alert messages when there were ambiguities it could not resolve. The system specialists would then consult the author or editor or subject specialist to find a less ambiguous or clearer representation of the thoughts or information. The system specialists would also undertake to maintain and upgrade the knowledge corpus. They would probe it for statistically unexpected clusterings or basic abstract correlations that had not yet been detected. These might imply possible new knowledge structures and would be called to the

15 Ibid., Licklider, JCR., p. 25.

attention of researchers in the substantive fields but also researchers in the field of knowledge structures. System specialists would also make contribution to the teams of information scientists seeking continually to improve the representation language and processing of information into knowledge.

The substantive users would also contribute to the evolution of the procognitive system both implicitly and explicitly. Users would be expected to examine the results they receive to their queries or questions and refine their search prescriptions or questions. They would indicate which results they find most insightful by choosing to use some over others. The system's programming code would be open and users would be encouraged, if they wanted, to make suggestions of improvements to the representation language. Licklider expected that substantive users would contribute significantly to the development and improvement of the procognitive system. The system would encourage human-human interaction, group use and easy methods as part of the system to get to other users, to system specialists or to librarians when human help is needed. The procognitive system would be programmed to utilize such user action as feedback and adapt itself toward the goal of improving future results. Licklider conceived of the procognitive system as a self-organizing and adaptive 3-way partnership or symbiosis of humans, computer systems and the corpus of knowledge. Each was expected via feedback and adaptation to change and grow. The fundamental purpose of the procognitive system would be to improve the usefulness and promote the use of the body of knowledge so that human purposes were rewarded with greater success.[16]

16 Licklider scaled his vision of the procognitive system from his experience in the early 1960s. His experimental system was only big enough to hold three documents. In the 70s and 80s other researchers made progress dealing with databases of abstracts and later of "paragraphs and chapters, tables and pictures, abstracts ... references, reviews and notes, catalogs and thesauri." Small scale prototypes of procognitive processing appeared in the 1980s. By the mid 1990s it was possible to use supercomputers to test prototype semantic-like representation language processing of large databases. In one such experiment, the Medline medical abstracts database was processed. The Medline database consisted then of about 9.3 million medical text abstracts. This corpus was processed using a generic noun phrase extractor set of programs. The process yielded over 270 million noun phrases correlated with term co-occurrence frequencies. The 45 million unique phrases where indexed to the abstracts that contained them. A concept space was created as the knowledge corpus testbed for medical queries and searches. Physician collaborators were given access via a web interface to the research prototype system. Their reaction was reported as "highly positive". Anecdotal evidence was given that searching in the concept space was far more useful and much quicker than searching in human coded indexes. The researches who were doing this work saw it as a beginning prototype implementation "far more semantic than syntactic" of the kind Licklider envisioned. See, Schatz, B., Information Retrieval in Digital Libraries: Bringing Search to the Net. – In: Science (Washington, DC). 275(1997)17. p. 327 – 334. Online at http://www.canis.uiuc.edu/archive/papers/science-irdl-journal.pdf.

Licklider's procognitive system would process the whole corpus of recorded thought and information in order to capture the semantic relations and content within the data across all discipline lines. Licklider expected that the system could then be addressed and replied to in natural language format. The scholars and other users would receive natural language knowledge responses to their queries and searches. They would still however have to read and think and generate insights and make discoveries beyond what the system provides. The system would provide semantic-like concepts and answers but the humans would make the final and meaningful interpretation. Thus they could contribute back into the system in an ever-expanding symbiosis. Licklider projected that eventually humans would interact with the growing corpus of knowledge by controlling and monitoring the processing of information and requests into knowledge rather than by handling the details and all of the processing in their own brains. The processing in their own brains would then be doing the most advanced and creative knowledge work.

The success of the procognitive system Licklider envisioned depends upon one major expectation, the expectation that human-computer systems would be developed that could do highly automated and increasingly sophisticated semantic-like processing. This expectation includes the implication that significant natural language question and answer systems would also be possible. Licklider was writing in the mid 1960s when the field of Artificial Intelligence (AI) was in its promising infancy. Was Licklider like many of the people with whom he was working too optimistic about AI? Licklider explicitly explains that the success of the future procognitive systems would not depend upon breakthroughs in AI. He did not expect that the procognitive system needed "intelligent" contributions from computers. He wrote, "…useful information-processing services can be made available ... without programming computers to 'think' on their own."[17] Licklider had the intuition that semantic analysis and processing would be much more important then the syntactical research that was current in the 1960s. But he also felt that the line dividing syntactics from semantics may not be a sharp line. He suggested that as more subtle syntactical analyses were attempted and computers became more powerful, syntactic analyses may begin to show semantic aspects. Licklider had "no thought that syntactic analysis alone – whether by man or machine – is sufficient to provide a useful approximation to understanding."[18] On the other hand, he wondered, "…as subtler and subtler distinctions are made in the process now called syntactic analysis, [whether] that process will start to become semantic as well as syntactic."[19]

17 Ibid., p. 58 – 59.
18 Ibid., p. 131.

Licklider's intuition and vision was that syntactic processing would continue to increase in sophistication while hardware and network developments would likely make semantic-like knowledge processing possible. The research question Licklider left to be answered was what knowledge structures or forms or correlations or representations would prove must fruitful for the organization of the corpus of knowledge. For Licklider the library of the future was even more of a human-machine-knowledge symbiosis than Bush or Kemeny had envisioned. Licklider also raised the social/political question, would society set itself the goal of developing a procognitive system.

4. *The Google System, Syntactics and Semantics*

The visions of libraries of the future examined above were articulated from 1945 to 1965 and projected ahead to the year 2000. If we jump ahead to the beginning of the Twenty-First Century, the body of knowledge is being put more and more into digital form. That body is divided into at least two forms. There is the web page record accessible via browser and search engine of some billions of web pages of information. There is also a growing body of scholarly information processed into digital form by digital library projects or produced in digital form by publishers. Some of this body is in web form but much of it is in databases that are not reached by search engines. This divide will close as more digital library resources become available to search engine indexing systems.[20] The most popular method in 2004 for scholarly interaction with the corpus of knowledge available on the web is the Google, Inc. system. Even some scientists report more relevant and useful hits using the Google search engine than they find in specialized scientific search programs.[21] An article in Science traces the technology that is the foundation for such search engines as Google directly to the work of Licklider in the 1960s.[22]

The Google search engine was developed by graduate students as an open system.[23] Violating the original public essence of the Google project, the US Na-

19 Ibid., p. 141.

20 Young, J., Libraries Try to Widen Google's Eyes. – In: The Chronicle of Higher Education (Washington, DC). L (2004) 37. A1, A31 – A32.

21 Arms, W., Automated Digital Libraries: How Effectively Can Computers Be Used for the Skilled Tasks of Professional Librarianship? – In: D-Lib Magazine (Reston, Va). 6(2000)7/8. Online at http://www.dlib.org/july00/arms/07arms.html.

22 Schatz,B., Ibid note 16.

23 Sergy, B. / Page, L., The Anatomy of a Large Scale Hypertextual Web Search Engine. – In: Proc. The 7th International WWW Conference (Brisbane, Australia). 1998.

tional Science Foundation encouraged the graduate students to make their work proprietary. The current secret nature of the Goggle system and its for-profit purpose brings Google, Inc. into conflict with the open essence of the Internet, Usenet and the procognitive system envisioned by Licklider. Still the success of this search engine raises a question related to Licklider's intuition about syntactic and semantic processing.

The Google "web crawlers" are data analysis programs that download into a database and process upwards of a billion or more web pages every few weeks. They gather the words on each page (except for junk words) and make inverse indexes attaching to each word the URL of the web pages where it appears. They keep track of the position in the text where each word appears. They also index the URLs according to how frequently they are linked to and from other pages, giving greater weight to links from higher ranking pages. This indexing of the URLs requires processing matrices of the order of a billion times a billion. But Google's algorithms and computers perform these calculations routinely. The Google system also gives weight to font size and other formatting details. None of Google's processing is semantic. There is no intelligence in Google's indexes. Yet most users find the Google system powerful in quickly finding for them and ordering with a fair degree of relevancy web page sources that meet their search criteria.

Now envision as Licklider did if thesauri were generated which linked to each word in a search engine index other words related to it as synonyms or as equivalents from other fields of study and other relations. Envision if the words were linked to noun phrase and term switching databases, if statistics of term co-occurrence and density and clusterings were added for each page. Then the word and phrase and natural language queries and searches could draw all at once on these factors. Might we then be getting closer to matching concepts in the user's brain with concepts in the web page record? And envision what would result if we added to the web page record all possible databases and processed images and sound tracks. Would that not be closer to the semantic-like interaction with the whole corpus of knowledge at the heart of the procognitive system?[24]

24 Schatz, B. wrote in 1997, "By 2010, the vision will be realized with concept search enabling semantic retrieval across large collections. ... Information retrieval in the next century will be far more semantic than syntactic, searching concepts rather than words." Ibid., note 16. p. 327.

5. *Conclusion*

The visions from 1945 to 1965 suggested above resulted from the question of how to collect and organize and process the scholarly record so that it would be more accessible and attractive for the accomplishment of scientific and scholarly work. Bush and Kemeny and Licklider were technology enthusiasts who foresaw that the essence of a library, its knowledge content, need not be located in books or buildings. They shared a sense of the value of access to the whole corpus. They set the high goal for library and computer and knowledge scientists of developing a single human-machine-knowledge system that would make the body of knowledge more useful and accessible. There has been in the last 15 years a vast effort at digital libraries research. Some of this research has adopted this goal. Might the human-machine-knowledge systems like Bush's Memex, Kemeny's National Research Library and Licklider's Procognitive System still serve as a grand vision that will inform more digital libraries research and eventually lead to the enhancement of human life by giving all people a chance to benefit from intimate contact with the whole body of knowledge?[25]

25 The author wants to thank Prof. Dr. Klaus Fuchs-Kittowski for encouraging the preparation of this presentation and Prof. Dr. Walther Umstätter and PD Dr. Heinrich Parthey for inviting him to the „Wissensmanagement in der Wissenschaft" conference where it was first presented. The author also thanks Marcello Farabegoli of Universität Potsdam and Ronda Hauben of Columbia University for conversations which helped him work out his understandings. Also many thanks to the staff of the Inter Library Loan Department of Columbia University Library.

Gesellschaft für
Wissenschaftsforschung

GEWIF
BERLIN

Klaus Fuchs-Kittowski,
Walther Umstätter
Roland Wagner-Döbler (Hrsg.)

Wissensmanagement in der Wissenschaft

Wissenschaftsforschung
Jahrbuch 2004

Mit Beiträgen von:
Wladimir Bodrow • Klaus Fuchs-Kittowski Jay Hauben • Matthias Kölbel • Peter Mambrey • Erhard Nullmeier • Walther Umstätter • Rose Vogel • Sven Wippermann

Wissenschaftsforschung
Jahrbuch 2004

Deutsche Nationalbibliothek
Wissensmanagement in der Wissenschaft:
Wissenschaftsforschung Jahrbuch 2004 /
Klaus Fuchs-Kittowski; Walther Umstätter;
Roland Wagner-Döbler (Hrsg.). -
Berlin: Gesellschaft für Wissenschaftsfor-
schung 2011.
ISBN: 978-3-934682-58-0

2. Auflage 2011
Gesellschaft für Wissenschaftsforschung
c/o Institut für Bibliotheks- und
Informationswissenschaftwissenschaft
der Humboldt-Universität zu Berlin
Unter den Linden 6, D-10099 Berlin
http://www.wissenschaftsforschung.de
Redaktionsschluss: 15. März 2011