A world obsessed with searching for words:

Why is a minimalistic text based interface a cutting edge technology of today?[*]

Philipp von Hilgers[**]

Presentation at Media in Transition 5: Creativity, Ownership and Collaboration in the Digital Age. Massachusetts Institute of Technology 2007
4-270 Saturday, April 28 1:30-3, Call Session 5**,** Panel "Graphics, Typography, Interface"

1. Introduction

There might persons socialized by personal computers running Unix, DOS or Linux as their operating systems, who still know what a command shell is. Typically it is a black screen providing space for a single line of commands. If we compare the interface of these good old days to an interface which is among the most frequently used today, we will find that things have changed form a human computer interaction perspective: the screen is no longer black but white, and above the command shell there is a brand name: "Google".
As rumor has it, Google founders Larry Page and Sergey Brin asked usability expert Jacob Nielson what they should do with their webpage, the advice was, "Do nothing." In the course of this paper, I will argue that the interfaces of the common search engines might stay as simple as they are but they should definitely be extended by eye tracking techniques.

2. What it means to google

Just the advent of a new technology not necessarily makes a huge difference – not until it evokes new kinds of techniques and practices. That along with the rise of Google's search

---

* A great source of inspiration for this paper was a meeting with Thomas E. Hutchinson and Amy N. Langville at the College of Charleston earlier this months. Hutchinson, among many other research interests, is a pioneer in developing eye tracking devices. His expertise in applying eye tracking system in field where help is most urgent – among people with disabilities – caused me to look at human and computer from a different perspective. And Langville seems not only to known all the algebra of search engine ranking in detail but can also explain its principles in mathematical terms easiest to understand. Also I owe Michael Schiessl from eye-square.com many helpful insights in psychological aspects using eye tracking technology.
** Max Planck Institute for History of Science, Berlin, Germany and currently Visiting Scholar at the STS at MIT (p.v.hilgers@hu-berlin.de)

engine a new practice in using media was introduced is evident by the sheer fact that the verb "to google" has become a household word.[1]

So what does "gooling" mean, if even the company who introduced the brand name from which the verb derives is fearing that the word could encompass more than they offer as a service? To understand what "to google" implies at least two levels are important to distinguish: the user domain and the server domain. The user domain again can be differentiated into writing and reading practices.

The technology of search engines created something new by extending and changing the basic cultural technique of reading and writing. Most of the search words entered at search site reflect and anticipate somehow the response features of search engines. Search words don't specify so much a certain information as they exclude billions of information. It seems that users of search engine are much more determined in what they are not interested in than what the information should be.[2] More likely to google means to frame some sort of textual contingency. Googling is tickling the mind by forcing the possibilities of being surprised. Search engines can be seen as machines maintaining desires to be on a high level partly driven by the subject's catch words and partly by the server's response.

If with googeling a new pattern of writing emerged, what is there to say about reading the output of search engines? In general nothing can be learned instantaneously from the results of the search engines. The results inform about information. I guess a pretty common googeling experience is that the results do not bring up the desired information but at least some better search words. After a while of an iterative search session, it not absolutely uncommon that the initial search target gets totally out of focus. Search engines do not just help their subjects to fulfill a special intention, but they turn their user into an agency which keeps search engines running.

Let's turn now to the server domain, which requires opening the perspective from a individual stand point to a view including collective structures.

---

1 Wikipedia reports that the verb 'google' "was officially added to the Oxford English Dictionary (OED) on June 15, 2006 and to the 11th edition of the Merriam-Webster Collegiate Dictionary in July 2006." However, the company fears that a ubiquitous presence of the verb "to google" could "genericide" the brand name in ways which lie beyond the service the company want to be known for.

2 Fore example Ggoogling with the phrase „Jaguar Cars 2007" is to say: I am not interested in information about the animal Jaguar, not in classic cars, not in used cars, not in other cars than Jaguars.

The history of search engines had taken a remarkable twist when two graduate students took a leave of absence from their computer science Ph.D. program in Stanford and started the company Google in 1999.

As often can be observed in the history of technology, Page and Brin did not simply develop a new technique but ignored radically already existing techniques. Before Page and Brin entered the scene, the art of search engines development was focused for decades on the semantical level of information. Hope was high that information retrieval creates an artificial model of text understanding. In fact, even in the end of the 90's search engines still had difficulties to rate somehow the information of internet recourses.

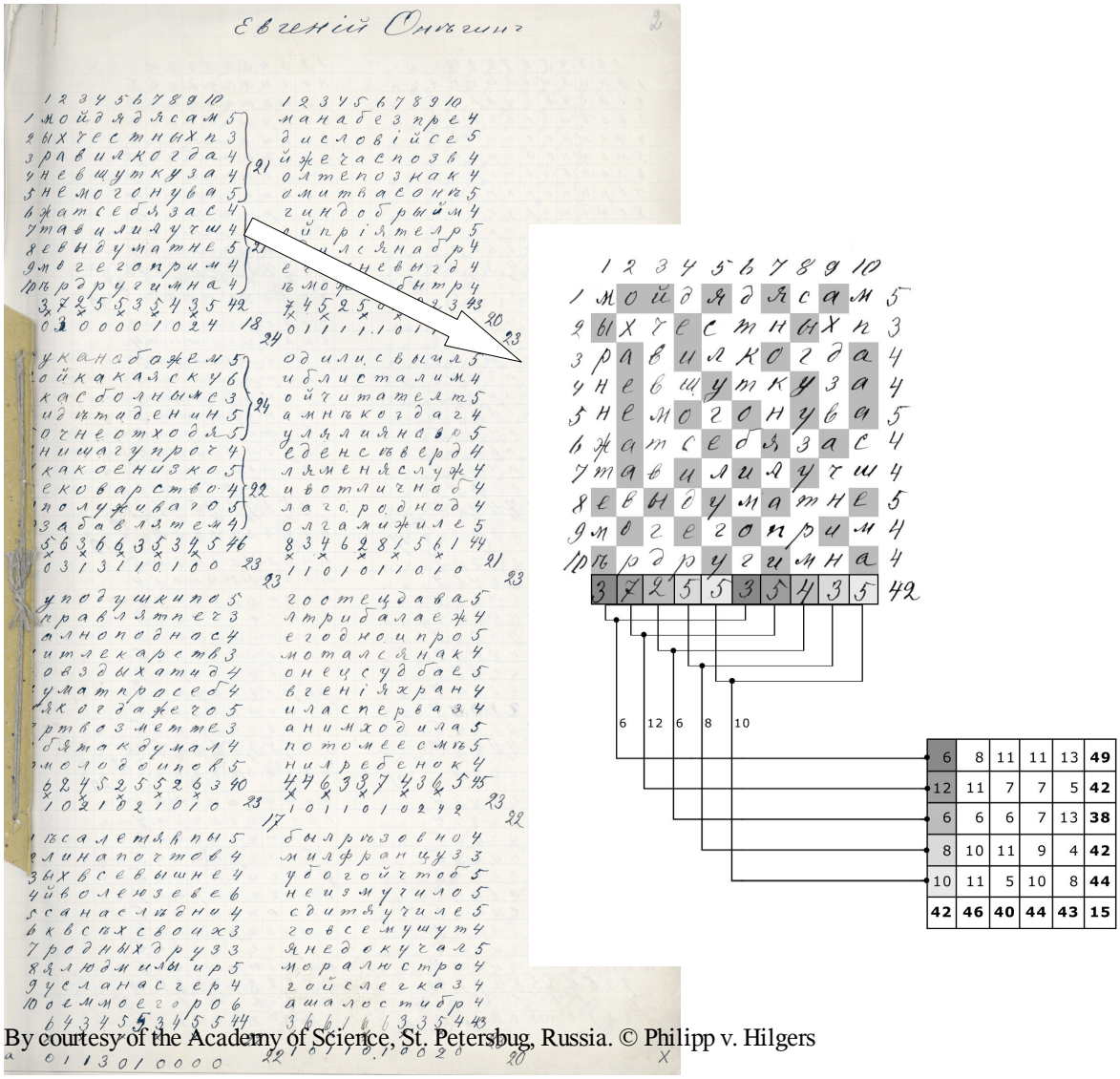
The Google approach left this developments behind by making use of the fact that referring to webpages by links can be taken as sign of appreciation. Google's PageRank ignores the content and utilizes the human capabilities to highlight certain information with ease. The PageRank's algorithm measures the quality of a webpage by analyzing how many and the way other pages are referring to it. The credibility and quality of the referring webpages are likewise measured by the way they themselves are the subject of the links of other webpages. So we can speak of a circular definition of the webpage's quality.

In Page's and Brin's eyes, the web surfer is more or less on a random walk through the internet. Entering a search word is manifesting his current point of interest in the internet, and Google's response can be compared to advice saying if you are interest in such a topic, most web surfer in the same situation went to this or that webpage. So what Google algorithm does is to predict the transition probability going from one webpage to an other based on how many signs point to a particular web side.

The mathematical model behind the PageRank algorithm was originally developed by Russian Mathematician Andrej A. Markov in the beginning of the 20th Century. Markov applied his so-called Markov Chains to predict the probabilities of combination of vowels and consonants following each other in Alexander Pushkin's poem Eugene Onegin. (fig. 1) With Markov chains, not only a model was introduced which has shown to be useful for many application including page ranking, but also the idea was born that language is not only ruled by lexical, grammatical and semantical conventions, but also by a stochastic

process.[3] (Fig. 2)

3 See Philipp von Hilgers and Amy N. Langville, The Five Greatest Applications of Markov Chains. In: MAM 2006: Markov Anniversary Meeting. Ed. Amy N. Langville and William J. Stewart. Raleigh 2006, 155-168



By courtesy of the Academy of Science, St. Petersbug, Russia. © Philipp v. Hilgers

Fig. 1 The first 800 letters of 20,000 total letters compiled by Markov and taken from Alexander Pushkin's poem "Eugene Onegin." Markov encoded the letters into a numeric matrix and evaluated the distribution of the chains of vowels and consonants.

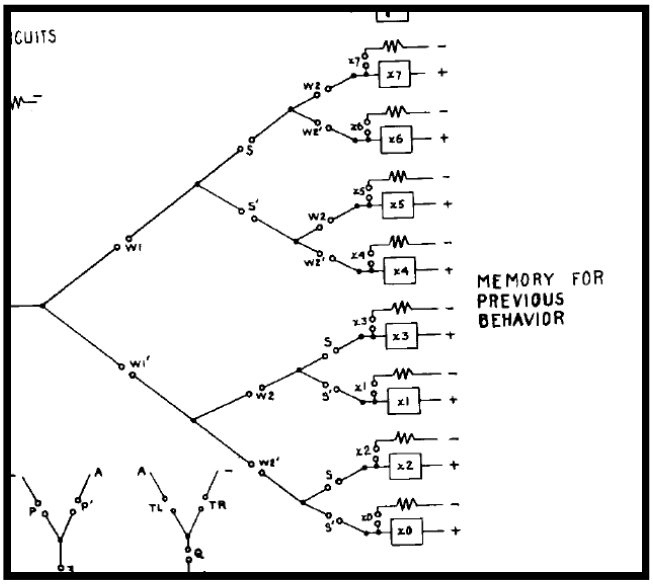


Fig. 2 Part of a circuit diagram of Claude Shannon's Mind Reading (?) Machine from 1953. The MRM is the first hardwired implementation of a Markov Model. It plays the game "Odd and Even" against a human opponent. In a more general view, the MRM was also comprehendas pattern learning machine.

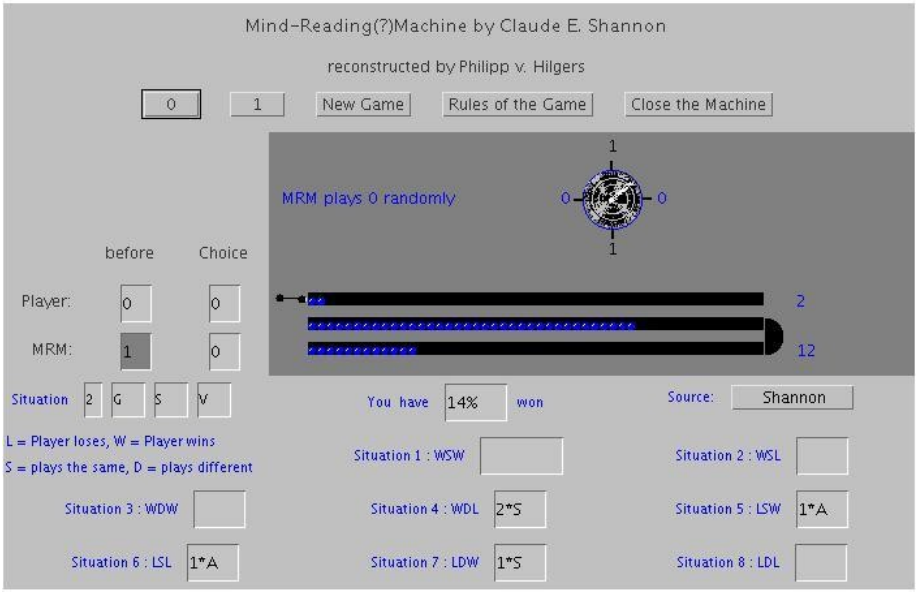


Fig 3. A reconstructed version of the Mind-Reading (?) Machine by the author can be found at: vlp.mpiwg-berlin.mpg.de/exp/hilgers/mrm3.html

3. The Missing Link between Search Engine Ranking and Eye Tracking

Even though the Google PageRank is based on a powerful mathematical model, some people are criticizing its effects. Among other reasons, they fear that Google is enforcing an opportunistic course of action: Webpages with many links pointing to them are highlighted by search engines and highlighted webpages are more probable candidates for being referred to than such which are not, so already quite visible pages might even climb up to a still higher ranking position. However, a high ranking position not necessarily reflects what a webpage actually has to offer. In some cases it is disturbing that an information is perpetuated by some sort of algorithm longer than necessary just because the algorithm has no direct access to the implications of the information it rates.

It takes some time to code a webpage to refer to another webpage, and it needs a reason to do so. While not everybody has the understanding and the facilities to modify webpages, companies can and have manipulated Google ranking systems on the large scale with so called linkfarms. On the other hand the wiki principle has successfully shown how the barrier of involvement in internet activities can be decreased. The success of internet encyclopedia Wikipedia is mainly based on the wiki principle.[4]

However, the reason why the Google PageRanking is based on links as a indicator for preferences is a pure technical one: they are so far the only computable accessible indicators. The other option is to let people explicitly judge how useful a webpage is. It seems that Wikipedia's for-profit search engine project "Wikia search" plans to strengthen the web surfers part in ranking web sides.

While Google will certainly not extend their search engine with methods relying on any kind of cooperation by their users and/or providing a transparency which allows others to manipulate or modify their ranking system for better or worse. In consequence, Google provides probably the most advanced search engine for the internet, but it might not be the best possible one. The simple question is: Are there better indicator than links to compute the attraction of a webpage, an indicator which Goolge's approach can't take into account? I think

4 Of course, one may argue that a greater effort to link one webpage with another reflects somehow a clear decision while a much more easier way of showing interest in webpages opens the door for arbitrary or mindless manipulations.

there are better indicators, but they are beyond Google's reach – because they exceed the scope of the internet. These indicators can be found in the user space.

The user space is a domain where the internet ultimately ends.[5] It is the space where real persons are sitting in front of computer displays, key boards and mouse pointers.[6] The user space is the place of human computer interactions.

It is in the user space where people will show the first reactions to some sort of information. They are reading the information and they are reacting to it. For most of the time the only indication that a user's attention is called by a piece of information are his or her eyes' movements.

Instead of evaluating the importance of a webpage by measuring its somehow artificially interconnectedness to other pages, its grade of attraction could be defined by the amount of eyeballs turned to it. To take eye gazes as a basis of evaluating the relevance of a webpage means to shorten dramatically the time needed until significant links are present for indexing. Up to now, it is indefinite when and if a webpage will show up and make a reference to another webpage. With eye gazes, pointers of interest appear immediately.[7] It turns webpage ranking from an off-line system into a real time system.[8] Thus, the marriage of eye tracking and information retrieval seems to be a perfect fit, as they could create a feed back loop in real time, where both systems respond in the fastest way they can.

Still, there are some serious hindrances. Even the capturing of eye movements at nearly any desirable level of accuracy is technically solved, the costs for the technique are still to high to address the mass market. But the costs are falling.[9]

---

5 The user space, as I understand it, includes all data of a personal computer which cannot necessarily be reached via internet. In some sense Google's strategy is to push even the user space deeper into the internet by offering online office applications.

6 It is probably necessary to remember that this space still exists, since people absorbed by their computer displays tend to overlook their nearest surroundings and perceive the internet as a pure virtual space.

7 Unlike eye movements, the perception of some kind of information is not necessarily followed by mouse clicks or movements and keyboard strokes and if they are not as fast the reactions of the eyes.

8 Google's PageRank is already known for the advantage that once linking structures in the internet are analyzed, it can response instantly to any key word. In contrast Jon Klein otherwise excellent page rank approach Hypertext Induced Topic Selection (HITS) needs to compute the rank in a time-consuming manner for every search word, at least in its original version. See Amy N. Langville and Carl D. Meyer. Google's PageRank and Beyond: The Science of Search Engine Rankings. (Princeton: Princeton University Press, 2006), 30 and Chapter 11

9 The most expensive part of an eye tracker is the digital camera. The costs of high resolution infrared sensitive cameras are decreasing rapidly as the market for surveillance cameras exploded in the last years – needless to say why. Manu Kumar from HCI Group in Stanford already raised the question of when

Still, even if eye gaze tracking device would become as common as sound cards are already today, what can eye tracking data reveal? Intuitively we think the human gaze indicates the orientation of attention, and for this reason also reflects interests. However, more specific investigations of eye movements show them to be often ambiguous in terms of representing what human subjects have in mind.

Nevertheless, I will present some first concepts how gaze-augmented interfaces can overcome the complexities which are associated with eye movements. In this regard it is not unimportant to recall some physiological basics of vision. In general, the field of vision takes up 200 degrees of range, while the fovea, the central part of the retina, only covers about 2 degrees. The fovea is responsible for the greatest acuity in vision. For example, due to the fovea we can read letters. While reading a book, generally no more then nine letters fall onto the retina. Letters which are not seen by the fovea can hardly be recognized. Obviously that is the reason why we have to turn our eye to some kind of visual object in order to investigate it. (Fig. 5)

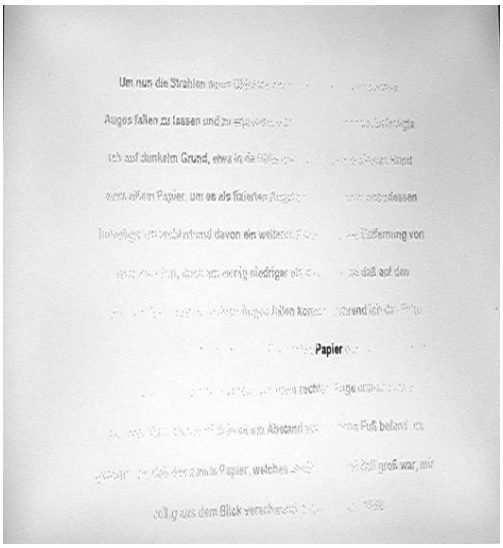


Fig. 5 Still from a screen recording where a text is blurred except from a test person's fixation area. Since a eye tracking system ensures that the mask of clear view moves synchronously with the user's eye gaze the test person is able to read the text while he is adopted to the system. See the whole video at: vlp.mpiwg-berlin.mpg.de/exp/hilgers/benjamin.mov

notebooks with a camera already embedded will be extended by an eye tracker mode. See Manu Kumar, Reducing the Cost of Eye Tracking Systems. Stanford Tech Report CSTR 2006-08, April 2006

Now, let's say two words are shown somewhere in the left periphery of a display. (fig. 6) If we turn our eye to one of the words, then the eye movement seems just to fulfill a necessary condition to read the word. Having the word in the center of the field of vision we created a sufficient condition to read it. If this model holds, then turning the eye to this particularly word and not to the other means in no way that we are interested in this word more than in the other. So we are probably turning our eye to one of the two words just to recognize it better or at all. Interactive eye gaze based interfaces run into a problem if they are valuing the fixation on the word as a selection. Before a user even has time to think about the meaning the word he is reading, it becomes selected.[10] Does it seems really convincing that turning the eyes to one of the words has to be an arbitrary choice since none of them are fully recognized yet? While reading a newspaper, are we this little ant that moves from header to header, from word to word, ignoring everything besides a tiny area of interest? Thus, not less plausible is the understanding that a preprocessing perception mode directs the eye to one of the words based on non random factors. Such a preprocessing mode must be assumed to be an unconscious activity, as it happens in a time span too short for being intentionally performed.[11] It is the latter strategy which should be amplified by an gaze-augmented interface in such a way that the eyes are not so much directed by an intention but by textual information attracting the eyes and orchestrating their movements. What the interface should do is to short cut the scan path of the gaze to a minimum. An optimal solution would make such eye movements unnecessary which are setting the stage for the recognition of an area with information. Instead the interface itself ensures that the area with some information is always recognizable and is covered perfectly with the user window of attention. This area is, then, fed with a stream of information. All eye movements still showing up should indicate

10 The problem is named “Midas touch” among interface experts after the Greek king who is said to have turned everything he touched into gold. However, human computer experts don't find any gold but discomfort with interfaces acting in this way. Robert J.K. Jacob brought up the “Midas Touch” problem while investigating interaction techniques which could incorporate eye movements into the user-computer dialogue in a convenient and natural way. See Robert J. K. Jacob, Eye tracking in advanced interface design. In: Woodrow Barfield und Thomas A. Furness (Eds.), Virtual Environments and Advanced Interface Design. (Oxford: Oxford University, Press, 1995), 258-288.

11 Findings in the field of cognitive psychology confirm evidence for both models: In a bottom down strategy, perceptional tasks are carried out in a general manner ignoring what distinguishes a reading scenario from any other. This strategy is said to compete with a bottom up strategy which depends on the particular bits oft information.

only which part of information is more attractive than another. As soon as such a more or less pre-conscious decision for looking closer at a catchword is detected, the content of the area is stepwise refined. Providing more and more accurate information can be done by state of the art search engine rankings, while the data used as the basis can shift from indexed links to a history of the paths of gazes.

The principle to let the interface carry the information in the center of the user's focus is quite straightforward.[12]

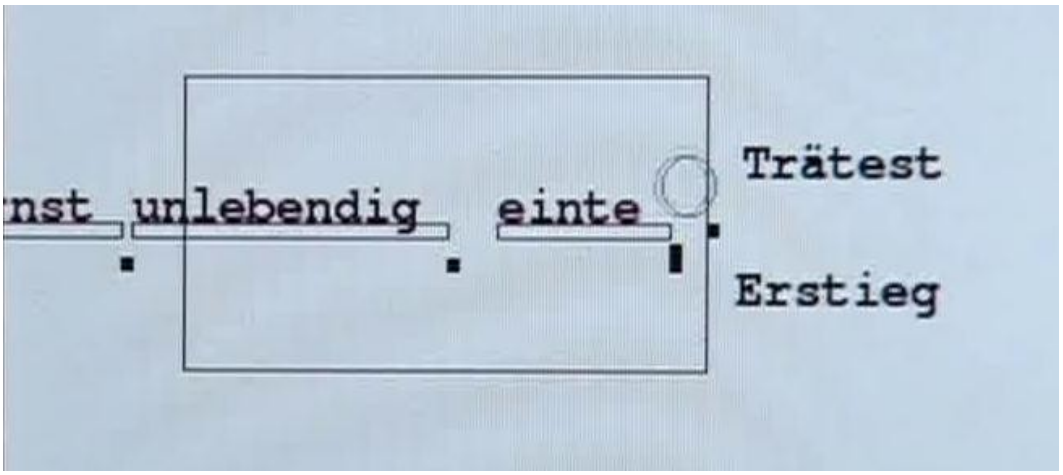


Fig. 6 Still from a screen recording. The test person is reading a yet unfinished sentence. The area of his fixation is marked by a circle. In the moment his gaze turns to one of the two words to the left the two words will disappear. In the most cases, the test person will then return looking at the end of the sentence. At this moment the sentence becomes updated by one of two words before placed at the left side. The word which attracted the test person's gaze first will become part of sentence. The word to choose are provided by a markov generator.
See the whole video at: vlp.mpiwg-berlin.mpg.de/exp/hilgers/reader.mpeg

12 In fact, implementing such a principle will bring up a dilemma: As soon as items such as words are animated and moving on the display, the human observer tends to react very strongly to those items for the sheer reason of being moved around. The eye movements, in such a case, do not reveal any higher cognitive processes but just low level reflexes. Still, there are ways to escape the dilemma. One is fading information smoothly in and out. An even more elegant solution can take advantage of the phenomenon that during rapid eye movements, so called saccades, the human vision system is suppressed down to a level that a user is not able to notice any kind of visual appearance. During saccadic suppression words can be exchanged in a way that the user reacts to only the new information without being sidetracked by the process of updating the visual content of the display. Keith Rayner introduced such a technique for an experimental design testing parafoveal preview benefits. See Keith Rayner, The perceptual span and peripheral cues in reading. Cognitive Psychology, 7 (1975), 65-81.

More information about Hendrik Schumacher's and Sebastian Kaiser's ryhming markov generator can be found at: www.versquelle.de

4. Outlook

Coming to the end of my paper I would like to sum up that the developing of pretty cool media is in reach. As is generally known, Marshall McLuhan distinguished between cool and hot media. Cool media, as we can interpret McLuhan, are cognitive ambitious media. They operate on the highest abstract level with a low stream of symbols occupying not so much our senses but our cognitive skills. (fig. 7) Hot media are in contrast media flooding our senses with data in order to simulate scenarios already existing. While hot media tend to produce redundancies, cool media should always try to bring us to the limit of our cognitives gifts with a minimum of visual data. Last but not least, taking eye tracking data for a ranking basis of internet resources, a crucial ratio can be changed. Let's call that ratio the seeking/linking ratio. If you are a frequent user of the internet, ask yourself how many links you add to the internet per week and how many search words you enter at search sites during the same time. The group of those who search the web and the group of those who add links to it should converge. By sharing gaze points, you can learn from others the pattern of visual recognition to comprehend abstract data, and others can learn your way of reading the world. And it will lower the entry level for people who are still out of the game. Eye tracking devices have already become helpful devices for people who are suffering from being paralyzed for different reasons.[13] These people belong to the media avant-garde and live already in a presence of media involvements which might be our future. As we can learn from media history, people with some sort of handicap are the early adopters. Both Alexander Graham Bell, the inventor of telephone, and Pastor Malling Hansen, the inventor of the first produced typewriter, came to their technical solutions working as teachers of the deaf and dumb.[14]

13 Also it helps to diagnose difficult mental conditions caused by autism or schizophrenia years before they become evident.

14 Bell's invention of the telephone was in some way a continuation of the project of his father's, who created a “Visible speech” for deaf-mutes. Likewise Hansen hoped that his typerwriter to make his deaf mute pupils “speak with their fingers”.

Among the first people who bought Hansen's typewriter was German philosopher Friedrich Nietzsche. With the sentence he hammered with his typewriter on paper I would love to end my presentation: Nietzsche wrote: “Our writing utensils collaborate in producing our thoughts!”

Fig. 7 Still from a screen recording showing the author's Discourse Analyzing Machine (DAM) in action. The blue circle marks the gaze point of the user. The user just “selected” with his gaze the subject category “Activities” from a cluster of subject categories. All subject categories are moving to the periphery of the display stimulating the user to have a look at one of the words. The user will hear the selected word spoken out loud by a text-to-speech engine. The subject categories are derived from the hypertext version of an old dream book from Byzantine. They lead to a sentence interpreting a dream. Have a look at the whole video of the screen recording: vlp.mpiwg-berlin.mpg.de/exp/hilgers/dreambook.mpg
Florian Cramer's hypertext version of the dream book can be found at: http://www.thing.de/projekte/7:9%23/dream_book.html