THE WEB—EARLY VISIONS, PRESENT REALITY, GRANDER FUTURE

John McCarthy, Stanford University
mccarthy@stanford.edu
http:/wWww-formal.stanford.edu/jmc/
October 22, 2004

- Turing—1936 Universal computer

- Vannevar Bush—1945-Memex

- McCarthy—1961—Time-shared Computer Utility, motivated by Advice Taker proposal

- Licklider—1961—Man-Computer Symbiosis

- Roberts—ARPAnet → Internet

- Engelbart—197x—Mouse, linked documents

• Berners-Lee—199x—World Wide Web

• Brin and Page—199x—Google—first adequate search engine

• other prophets—Nelson, etc. whom I neglect undeservedly from ignorance.

SOME EARLY PREDICTIONS WORKED OUT—OTHERS NOT

• Time-shared public utilities. Modest success. Lack of machine power, needed too much hand-holding. Worked fine in labs

• Stanford AI Lab news service, 1972. Prototype web newspaper.

• Access to all the world's books. Still hasn't happened. Making steady progress for scientific articles. No economic model for literature except what's out of copyright. John Ockerbloom at the University of Pennsylvania links to more than 20,000 free-to-read books.

• On-line buying and selling. I didn't predict Ebay auctions.

TIME-SHARING vs. PERSONAL COMPUTERS

• General purpose time-sharing advocated in 1959, realized in 1962. Gives each user a share of a computer at his fingertips.

• TX-2 about 1960 was a $500,000 personal computer. SUN was a $20,000 personal computer.

• Multics got obsessed with security to the extent of impracticality. IBM followed M.I.T. into the swamp with TSS but abandoned it.

• Unix got rid of impracticalities but still inherits some characteristics of very small machine.

• Best for its time was PDP-10, but it wasn't pushed hard in competition with IBM. Better operating system than Unix or Mac or PC.

• PCs were too weak initially, and the operating systems inherited some of the weaknesses and kept them in vastly stronger computers.

TIME-SHARING vs. PCs, part 2

• The PC hell is system administration. Need AI to do it properly on a mass scale.

• Software bloat

• Operating systems as products require the user to do sysadmin for every new version. A time-sharing subscriber would not. There are probably 100 times as many system administrators as would be needed.

• Presently promised "set top boxes" seem to aim at monopoly. Sysadmin centralized but probably too little and too late.

FUTURE

• The present web is pretty good. The users will do ok even without new ideas. Those whose business is new ideas will suffer unless they have good new ideas. The dot com crash was substantially due to a large number of bad or trivial new ideas.

• Everyone has trouble using something new. Systems must understand user states of confusion. Trivial example: a user confuses IP address, email address, and URL.

• It is more important for a system to understand a user's confusion than to offer sympathy. (Some advocates of "emotional computing" are hoping to get by with sympathy. My bet is that sympathy will only produce annoyance.)

1970 MODEL WORLD OF THE FUTURE

Here are some 1970 model new ideas compared with what happened.

This 1970 conference article was published in *Man and Computer,(Karger, Basel 1972)*. It's available as www.formal.stqnford.edu/jmc/hoter2.html.

Here's a fragment of what I wrote in 1970, with notes in blue.

****

At present, a newspaper or a book is a package produced by a large organization.

In our new system, the physical production disappears allowing a much smaller organization

to put out the same packages of text and pictures. Moreover, the user does not face a one shot decision to buy Life or Look. He will be able to read the 'cover' or table of contents of each and read such items as strike his fancy, and the system will bill him for what he reads from each source. In fact, since the cost of keeping a file of information in the computer and making it publicly available will be small, even a high school student could compete with the New Yorker if he could write well enough and if word of mouth and mention by reviewers brought him to public attention. What, then, is a publication in the new information system?

Note 2000: I didn't think of the resistance to being displaced these organizations would be able to mount. The clearest examples of such resisters are the publication organizations of scientific societies which are in principle non-profit organizations.

Note 2004: Four years later, the biologists have taken the lead in creating on-line journals that compete with print journals. Their financial basis is page charges, which works for science, because page charges are a small fraction of the cost of doing the research. It won't work for people who make a living by writing.

A publication is an organization that puts out a list of material it has edited and recommends to its readers. It helps its authors produce material that it thinks will suit the readers, and it has a financial arrangement with them about splitting the proceeds.

There can be a wide variety of publications of different standards of writing and editing and different budgets for carrying out these activities.

However, they will all be equally accessible to all readers, and the only justification for an expensive editorial organization will be that it can

produce a more popular package. The price of reading a package can be set by the publishers.

Note 2004: This ignores the copying problem. There still isn't a general purpose pay-by-the read mechanism.

A reader may feel that he needs help in finding his way through the totality of literature available to him. Various people will be eager to make a living by providing it. A bookstore or library is a program that when called shows the 'covers' of publications. Reviewers will produce lists for him and make money when he reads their lists or by kickbacks from the publishers. 'Reading advisers' under some catchier name will offer to generate lists just for him according to a profile of his interests.

Note 2004: This hasn't happened enough to make writers independent of publications. The

1970 article didn't take into account the importance of publicity.

Advertising in the sense of something that can force itself on the attention of a reader will disappear because it will be too easy to read via a program that screens out undesirable material.

Note 2004: This hasn't happened enought to discourage advertising. Also I neglected to predict spam. I am temperamentally an extreme optimists, but the pessimists didn't predict spam either.

Another effect is the possibility of frequent revisions of articles and books. An author can take into account new facts or other people's criticisms, and the revision will take effect immediately.

Note 2004: I do that with dated footnotes.

Public controversy can be carried out more expeditiously than at present. If I read something that seems controversial, I can ask the system if anyone has filed a reply. This, together with an author's ability to revise his original statement, will lead people to converge on considered positions more quickly than at present even if they do not come to actual agreement.

Note 2000: There are various proposals, but this hasn't happened yet. One can imagine Bush and McCain "truth squads" putting on their candidates' web sites arguments against the positions of the other guy. Personal attacks too.

Note 2000 June 1: Today's New York Times has an article entitled "E-Mail Messages to the Press Have Made the Gore-Bush Race a Cyberwar" recounting how the Gore and Bush campaigns send dozens of messages per day

to reporters. I suppose this is a partial realization of my 1970 prediction.

Note 2004: The campaigns have their web sites, but I think this still isn't the main placed undecided people go to see arguments refuting those of the other side.

Famous authors will not need publishers because their loyal readers will have the system find their stuff automatically.

Note 2004: A try at this failed because of copying

To summarize: the new information system will promote intellectual competition by reducing the price of entry, will permit readers to be selective, and will allow authors to revise material until they are satisfied that it withstands

criticism as well as it ever will. This should make intellectual life more interesting.

Note 2004: This doesn't seem to happen much. Instead of perfecting their earlier analyses, bloggers just bombard their opponents with new stuff.

The financial aspect of writing would presumably be as follows: a piece of written material has a price for reading it (this price may be zero for amateur writing, political propaganda, advertising, and for scientific journals). The reader's account is debited and the account to which the material belongs is automatically credited. The reader will have the system balk at what he considers overpriced material.

The new information system will have a profound effect on buying and selling. Sellers of

movies, groceries, automobiles, plumbing services and cures for baldness will find it advantageous to list their wares in the information system together with current prices and availability. The user can place an order through the system as he can by telephone, but he can do much more:

Note 2004 : This happened, but isn't revolutionory.

(1) He can call on someone's program to scan the sellers of sports cars and propose what it considers the best deal. This program might even negotiate with programs representing the sellers.There's some of it now.

(2) He can tell the system whether last year's cure for baldness worked and get a summary of the opinions of those who bothered to record

their opinions of the cure he contemplates trying now.

(3) He can make an airplane or hotel reservation by interacting with a program the airline or hotel reservation company has written to tell him what is available. He need not suffer the delays you now get when you call an airline or travel agent at peak hours.*

(4) Individual design and construction services can be offered through the system although this requires the development of computer-controlled manufacturing techniques for various types of articles. The idea is that automated design programs can produce designs for articles meeting individual specifications. Either by himself or in consultation with an expert, an individual would use the system to produce a design and display how it would look and possibly how

*All this has happened.

it would perform. Candidates for individual design include clothing, furniture, boats, electronic equipment, houses, and even cars. The system would then produce the instructions for controlling machine tools, fabric cutters, and also printed instructions for the hand parts of the operation. In general, it should be possible to make single objects at little more cost than present mass produced objects. In some cases, there would even be savings, because mass production requires estimates of demand that are often wrong resulting in inventories that are expensive to sell or even have to be sold at a loss; the cost of this is made up by a general increase in prices.

Note 2004: This hasn't happened yet. Maybe it will.

There are many more useful services that can be offered through the new information system

and again the system is conducive to competition. Writing and storing a program and announcing its availability can be a very low capital operation, and the system can collect whatever price has been set for its use.

Note 2004: In the world of pcs, this is far less convenient than in a world of time-sharing—or than it should be.

Note 2004: I greatly underestimated the important role refereeing and publicity of all kinds plays in creating reputation and getting attention to ones ideas. Example: Maybe I could have headed off some blunders of XML by publishing my *Common Business Communication Language* elsewhere than in in a one shot IBM conference proceedings.

LOW AND HIGH LEVEL HELP

Example: swindle protector

- Low level knows about specific swindles.

- Higher level can identify variants of the Nigerian scam.

- High level—knows facts about swindling in general.

Example: Understanding a user's confusion.

• Suppose the user confuses IP addresses and URLs. Suppose a program asks for an IP address, and the user gives a URL. Most present programs will simply put up an OK box that says "wrong format". The user may just worry about the format of the URL. A system designer who anticipated the confusion would have the program say "You gave me a URL when I asked for an IP address."

• More generally, system administration requires knowledge and reasoning. Evidence: The people who spend several hours fixing my problems obviously think a lot. They understand enough to fix my problems, but they don't understand enough about how they do it to automate their work.

FORMALIZATION OF CONTEXT

I can't give a conference speech without at least some logical formulas.

We write

$$c : p$$

to assert p while in the context c. Terms also can be written using contexts. $c : e$ is an expression e in the context c.

The main application of contexts as objects is to assert relations between the objects denoted by different expressions in different contexts. Thus we have

$$c : Does(Joe, a) = SpecializeActor(c, Joe) : a,$$

or, more generally,

$$SpecializesActor(c, c', Joe) \rightarrow c : Does(Joe, a)) = c' : a.$$

Such relations between expressions in different contexts allows using a situation calculus theory in which the actor is not explicitly represented in an outer context in which there is more than one actor.

We also need to express the relation between an external context in which we refer to the knowledge and awareness of AutoCar1 and AutoCar1's internal context in which it can use "I".

PHENOMENAL DATA MINING

• The data in a file of purchases in a supermarket, are a window on the phenomena, e.g. the customers and their demographics.

• My 1997 paper was well received, but there are no implemntations; they require commonsense reasoning using a common sense knowledge base as well as from the database being mined;
www.formal.stanford.edu/jmc/phenomenal.html.

• Unfortunately, making a commonsense knowledge base is difficult. Reasoning from natural language is even harder.

• 2000 challenge: Which editorials on the web advocated that Clinton be impeached, and which opposed it. Hint: To statistical document retrieval programs, "not" is a noise word.

COMMON BUSINESS COMMUNICATION LANGUAGE

My 1985 paper proposed a Lisp list notation for business communication. That notation is better than XML, but the proposaols went beyond notation to include propose standardizing elements of common business communications, e.g. offer-to-buy and delivery-method. To understand a complicated delivery request and combine it with external facts requires logical reasoning. See my www.formal.stanford.edu/jmc/cbcl.html.

BIG ADVANCES REQUIRE LOGICAL FORMULATION OF COMMON SENSE KNOWLEDGE AND REASONING

• Humans mainly communicate in facts, not just rules or programs.

• Humans reason to get new facts from old. Logicians formalized these rules Gödel proved them complete.

• Reasoning programs require full first order reasoning.

• Advanced help requires understanding the problem domain and usually understanding the user's state of mind.

SO WHAT'S THE GLORIOUS FUTURE OF THE WEB?

Programs that understand

- Substantial parts of natural language documents,

- Facts about the world,

- Facts about people's states of mind, including confused states of mind,

- can give good advice,

- and can put together programs from this information.