Dan Starer's comments on Information Technology

A computer background for Jeannie. All the experts I spoke to agreed that no one person could create a search engine like hers from scratch. These engines are written by large teams of computer scientists, most of whom have at least a masters degree in computer science or a "double E" (an advanced degree in electrical engineering). A psychologist, even one who minored in computer science in college, could not possibly create such an engine. Each graphical image such as an individual x-ray, ECG, or fingerprint requires many megabytes of space on a hard drive, optical disk, or other storage medium. The software involved to search such records is vastly more complicated and larger than the familiar software you and I might use for databases that contain text and numerical data. The programming language used to create or improve a search engine--C or C++--would not be known by someone without some computer science background.

At a minimum, Jeannie would have needed to minor in college in computer science and then either helped a professor with a project involving the scanning of medical images, or perhaps worked for a company in the field (or the FBI if you choose the wavelet route).

The computer language used to write Jeannie's search engine. While many of my comments regarding information technology from my three-page fax dated April 10 remain valid, I've now talked to a number of much more knowledgeable experts in several fields. Please note the suggestion about the search engine being written in 4PL or 5PL is *not* correct. Most likely such a search engine would be written in the programming language called C or C++.

Jeannie's breakthrough in pattern matching. A number of solutions were suggested to me. I've discarded the ones involving her creation or improvement of computer hardware. That would require her to have a heavy-duty engineering background.

On the software side: as mentioned above, it seems highly unlikely that Jeannie could design a completely new algorithm from scratch. People spend their lives studying algorithms, trying to improve the ways computers search and sort through things. This science is very leading edge. But Jeannie could certainly refine algorithms to make them more efficient, or use the hottest new types of algorithms in new ways.

Of the software solutions, two stand out. In both cases, scientists are just now applying these technologies to improving pattern matching among graphical images.

1). *Wavelets.* Since the FBI has chosen this technology for their own fingerprint analysis and pattern matching, this choice seems the most believable and realistic for Jeannie. From the articles I've supplied you can pick out one or more aspects of using wavelets that she improves. Perhaps she can be among the first to use wavelet technology to match patterns in dental x-rays (see the article *Image Indexing and Content Based Access to Databases of Medical Images with Wavelets).* Perhaps by doing so she refines her software such that it improves on the FBI's searching.

2). *Fractals, combined with Neural Neworks.* She might come up with something involving fractals. In geometry there is a concept that you can make big shapes based on repeating of smaller shapes. For example: if you cut a branch off a tree the shape of that branch is remarkably like the tree; cut off a limb from that branch and they too are remarkably similar; and the structure of a leaf from that limb looks much like the limb. Nature has many examples like this. Scientists create algorithms based on this fact of nature. They build software that looks at an image or a pattern and recognizes what is the repeating smaller pattern that makes the big pattern. That small segment is what the computer stores and manipulates. It is called fractal compression. Algorithms already exist to compress graphical images such as fingerprints. Jeannie could perhaps come up with a more efficient compression scheme. Even better, she might write an algorithm that speedily searches the already compressed data. One of today's burgeoning areas of research in pattern matching is figuring out more efficient ways to search compressed data. This type of algorithm would become even more efficient if it were combined with neural network technology.

Neural network software simulates the way human brain neurons interact to make decisions. This type of software has been found very effective in matching patterns, images, or fingerprints. The software teaches itself, through experience, variants in what it means to for something to match something else.

One of the experts I consulted is supposedly writing up a more detailed explanation of these technologies (at no charge), which I hope to fax to you shortly.

Whether you choose wavelets or fractals/neural networks, I can talk to experts quickly to answer your questions or to review anything you write about their use in *The Third Twin.*

Dental records. I spoke with a dentist/professor who is at the forefront of digitizing and computerizing dental x-rays. He agrees with the dentist I first spoke with that no present software matches dental x-rays and no database of dental records currently exists. This expert believes that any person in the dental profession reading *The Third Twin* will realize

such a software search of dental records is not possible today, or even in the fairly near future. This professor believes the problems of finding adult twins through x-rays are exacerbated by the likelihood that they would have lost or damaged different teeth through accidents and environmental factors

The professor did admit that theoretically you could do measurements of teeth or arch forms using dental x-rays from different people and determine that their measurements correlate to some degree. Perhaps if you mention teeth and arch forms it will lend more credibility your scenario.

(page 96) "I've devised a computer program"

(As mentioned elsewhere, software is written or designed, not devised.

"searches medical records and other databases for pairs."

(Two VERY MINOR issues:

1). DATABASES. The large amounts of relatively unorganized, unstructured data created when you digitize fingerprints, x-rays, etc., are usually called "collections" to distinguish them from traditional databases, which are limited to text and numbers, often in fields, or rows and columns. However, your readers will probably get confused if you use the word "collections" rather than databases.

2). PAIRS. Since the usual term for what her software does is "pattern 2matching," it might be more correct to say it searches for matches, rather than pairs.

(page 109) "There was no existing software to match pairs."

In a broad sense this is not true: many programs and algorithms exist for pattern matching. Perhaps you should be more specific here and say hers is the first software to search collections of medical images like ECGs to find identical twins.

"devised my own software" (Same as above)

(page 122) "Right now, the only truly reliable way to match fingerprints is with the human eye"

This is not true. As the enclosed articles reveal, the FBI can now do some highly successful pattern matching with their computers. Further, many of your readers will realize that security apparatus now exists that requires employees to let a machine scan their fingerprint before granting them access to a building. No human involvement is needed to match their print against the employers' database of fingerprints.

Also, on page 123, Ghita says "We'll run it at night, for minimal interference with normal use of the database." Does this indicate that computer searches of fingerprint files do work at the FBI or just that they serve to narrow the field? (Possible conflict with p. 122)

(page 239) "devised a search engine"

(page 300) The FBI has 200 million fingerprints? Best to check the figures given in the enclosed articles.

(page 508) "It takes an hour or so depending on the size of the database"

This makes no sense. If the database was small it might take much less than an hour. If the database was large it might take much longer. Factors influencing speed include the speed of the processor (or processors), the amount of memory available, the size of the cache, if one is present, and also software factors such as the type of operating system used.

Perhaps Jeannie can say the search has never taken more that an hour or half hour.

(page 508) "Does it interfere with normal data retrieval. No"

How could Jeannie know this for sure without knowing the operating systems, type of computer, amount of memory, and possibly other factors? A search of vast numbers of graphical images could well slow down the response time for other users.

I'd delete this sentence.

(520) Pentagon computer scene field 14 problem cited in earlier fax

It seems very unlikely that the same database would store huge, complex, unstructured medical images, along with addresses and names. I don't believe such an image would be stored in a field such as in a text/numerical database.

I've asked a computer expert for help on this and expect to fax a solution soon.

To answer the question in your fax, you could say the interface to access this database is Windows-based. But the computers used to find matches among ECGs would need to be mainframes or computers using lots of processing units working on the problem in parallel. Even a fast Pentium desktop computer could not conduct such complex

searches. Your readers should realize that a modem is attached and the mainframe or parallel processing computer in St. Louis is actually pattern matching the images.