SEARCHING THE WEB FOR CORROSION INTELLIGENCE

P.R. Roberge

Department of Chemistry and Chemical Engineering
Royal Military College, Kingston, Ontario, Canada, K7K 7B4

ABSTRACT

The rapid development of accessible computing power in the 1980s has led to the use of machine intelligence in every sphere of engineering. The incredible progress in computing power and availability has also created a tremendous wealth of information available at the touch of a few buttons. However, such wealth can easily provoke what is commonly described as 'information overload.' The massive number of connections produced by a single search of the web, for example, can greatly overwhelm users of this new technology. The rapidity of web searches is due to the synergy between progress made in network connectivity protocols, intelligent search strategies and supporting hardware. This paper will attempt to define the basic elements of machine intelligence in the context of corrosion engineering and examine what has been done or could be done to introduce artificial thinking into daily operations. This paper will also review some modern software systems commonly used for information processing and internet searches.

INTRODUCTION

The adequate transfer and reuse of information covering corrosion problems and solutions involves the development of information processing strategies that can become very complex. A typical corrosion engineering task requires handling different types of knowledge and disciplines such as metallurgy, chemistry, cost engineering, safety and risk analysis. The expected corrosion behavior of engineering materials is thus only one component of the multi-faceted life cycle management of systems. While the increasing availability of computerized information is making the topic of software accessibility and portability increasingly important, the main question that will be discussed here concerns the very nature of these tools in term of their intelligence or their user friendliness at introducing apparently simple solutions to complex problems.

THE INFORMATION AGE

The Early Days of Artificial Intelligence

The introduction of some sort of intelligence into machines is an old human quest that was given the name 'Artificial Intelligence' (AI) during a workshop sponsored by IBM in 1956. While the expression itself was intimidating to most people at the time, it also had an arrogant slant that, added to the poor performance of the early systems, encouraged a few detractors that started arguing in the defense of genuine human intelligence (Dreyfus *et al.*, 1986). During the decade that followed, researchers worked on developing intelligent systems on the belief that the intelligent behavior was primarily a function of reasoning techniques and that clever search algorithms could be devised to emulate human problem solving (Durkin, 1994).

A breakthrough out of the impasse created by this constraining belief came in 1965 when researchers at Stanford University were creating DENDRAL for the NASA. This computer program was to perform mass spectral data analysis for an unmanned mission the NASA was planning for a visit to distant Mars soil. The Stanford team realized that chemists performing these analyses often used rules-of-thumb to weed out patterns or structural features that unlikely accounted for understanding given data sets. The computer program that resulted of this effort could function equivalently to an expert chemist in recognizing molecular structures. This project was followed, at Stanford, by the development of MYCIN, a rule based expert system that used backward chaining and incorporated 500 rules for the diagnosis of infectious blood diseases. The MYCIN project was carried out in the mid 1970s and took approximately 20 person-years before completion. However, the success of that project was enough to rekindle a general interest in AI research and the production of AI tools.

During the 1970s, expert systems (ESs) were mostly a research curiosity. In 1985 approximately fifty systems had been actually deployed. One of the main attractions for developing ESs was the possibility to transfer some level of expertise to a less skilled workforce hopefully increasing its apparent level of expertise. Eventually the AI fever reached the corrosion community and led to quite a few important projects, as will briefly be reviewed in a subsequent section.

The Internet Revolution

The internet has revolutionized both the computer and communication worlds like nothing before. The invention of the telegraph, telephone, radio, and computer set the stage for this unprecedented integration of capabilities. The internet is at once a world-wide broadcasting capability, a mechanism for information dissemination, and a medium for collaboration and interaction between individuals and their computers without regard for geographic location. The internet represents one of the most successful examples of the benefits of sustained investment and commitment to research and development of information infrastructure. Beginning with the early research in packet switching, the government, industry and academia have been partners in evolving and deploying this exciting new technology.

The first recorded description of the social interactions that could be enabled through networking was a series of memos written by J.C.R. Licklider of MIT in August 1962 discussing his "Galactic Network" concept. He envisioned a globally interconnected set of [[_______]] through which everyone could quickly access data and programs from any site. In spirit, the concept was very much like the Internet of today. The combination of the powerful communication medium with other advances in computer interfaces and hypertext linkages had created the stage for the creation of a global environment that has revolutionize the history of modern computing. The timeline of important milestones in the history of Internet is presented in Figure 1.

The World Wide Web was released in 1991 by the European Laboratory for Particle Physics (or CERN), as a way for physicists to track each other's progress. The idea was that people working in different places could learn what each other was doing by looking at a hypertextual document set up on a computer which could be accessed through the Internet. This idea grew into the much bigger and large-scale operation that we now know as the Web. There are presently well over 4 million web sites, the computers which store and handle requests for web pages, and a great number of people all over the

Date		Operational Networks
1968	Formation of ARPANET working group (ARPA = Advanced Research Project Agency)	
1969	First 'packets' sent by Charley Kline at UCLA	
1970	ARPANET hosts start using Network Control Protocol (NCP)	
1971	BBN develops a terminal Interface Message Processor (IMP) or TIP that supports up to 64 hosts	15 nodes (23 hosts)
1972	The @ sign is chosen for its 'at' meaning	
1974	Design of a Transmission Control Program (TCP)	
1982	TCP and Internet Protocol (IP) are established as the protocol suite that is known as TCP/IP	
1983	First desktop workstations	113 nodes
1984	Moderated newsgroups introduced on USENET (mod.*)	> 1,000 hosts
1987	Email link established between Germany and China	> 10,000 hosts
1989	Creation of 'Archie' (archiver for ftp sites) by Peter Deutsch, McGill U., the first effort to index the Internet	> 100,000 hosts
	Development of a new protocol for information distribution (Tim Berners-Lee, CERN)	
1991	Development of 'gopher' the first friendly interface to the Internet	
	World-Wide-Web (WWW) released by CERN	
1992	The term 'surfing the net' is coined by Jean Armour Polly	> 1,000,000 hosts
1993	Mosaic, the graphical browser developed by Marc Andreessen (Netscape), takes the Internet by storm	600+ web sites
1995	A number of Net related companies go public, with Netscape leading the pack	> 25,000 web sites
1998	Release of Windows 98 with web browser integrated into desktop operating system	> 3,500,000 web sites

Fig. 1: Important milestones in the history of the Internet

world access the web for various reasons every day. The web is continually being enhanced and developed, due to rapid technological changes and the addressing of various questions and problems raised by the current state of the web.

A web browser is a software application used to locate and display web pages. Three of the most popular browsers are Netscape Navigator, Microsoft Internet Explorer (NCSA Mosaic), and Spyglass Mosaic. All of these are graphical browsers, which means that they can display graphics as well as text. In addition, most modern browsers can present multimedia information, including sound and video. A full gamut of tools has also been developed to navigate the web and search for specific information. The speed and functionality of these tools increase at a very fast rate. Table I shows a list of the most popular web search tools currently available.

One confusing issue is the difference between search engines and directories (Sullivan, 1998). Search Engines, also called 'spiders' or 'crawlers', constantly visit web sites on the Internet in order to create catalogs

Table 1

Name and web address of the most popular web search tools

Search Engines and Directories	**Address**
AltaVista	http://www.altavista.com/
Ask Jeeves	http://www.askjeeves.com/
AOL NetFind	http://www.aol.com/netfind/
Direct Hit	http://www.directhit.com/
Excite	http://www.excite.com/
Go	http://beta.go.com/
Google	http://www.google.com/
GoTo	http://www.goto.com/
HotBot	http://www.hotbot.com/
Inktomi	http://www.inktomi.com/
Infoseek	http://www.infoseek.com/
LookSmart	http://www.looksmart.com/
Lycos	http://www.lycos.com/
MSN (Microsoft)	http://www.msn.com/
Netscape	http://www.netscape.com/
Northern Light	http://www.northernlight.com/
Search.com	http://www.search.com/
Snap	http://www.snap.com/
WebCrawler	http://www.webcrawler.com/
Yahoo	http://www.yahoo.com/

of web pages. Because they run automatically and index so many web pages, search engines may often find information not listed in directories. Directories, unlike search engines, are created by humans. Sites must be submitted, then they are assigned to an appropriate category or categories. Because of the human role, directories can often provide better results than search engines. Yahoo is an example of a directory. Hybrid search engines also have an associated directory. However the repertoried sites have been reviewed or rated. Search engines have three major elements (Sullivan, 1998):

- A spider, also called the crawler. The spider visits a web page, reads it, and then follows links to other pages within the site. This is what it means when someone refers to a site being "spidered" or "crawled." The spider returns to the site on a regular basis to look for changes.
- An index. The index, sometimes called the catalog, contains a copy of every web page that the spider finds. If a web page changes, then this book is updated with new information.
- Search engine software. This program sifts through the millions of pages recorded in the index to find matches to a search and rank them in order of what it believes is most relevant.

Searching for Corrosion Information

The rapid development of internet tools for a community much broader than the group of scientists that initially created the medium is in itself an interesting challenge. Rarely before could one find new sources of information every month without even moving from the office chair. The main problem is not accessibility or availability any more but adaptability to constant change in the way information is presented and accessed. The functionality of the search engines and directories presented in Table II is constantly evolving in an exploding market. One important problem when searching the web is to stay focused on a search strategy since searches are offered at a discount only because they are subsidized by well advertised sponsors that can easily sidetrack a curious surfer.

In this diverse and very aggressive market, each search engine will rely on some fundamental offerings to attract a recurring fraction of web clients. This diversity is illustrated by the short description of the following search engines, the first two being highly popular, while the third engine, still in a beta version, searches the web based on popularity and traffic.

- ***Infoseek:*** Infoseek is one of the more popular search services on the web. It has a small-to-medium sized index, so it may not be the best place for those doing a comprehensive search of the web. However, it consistently provides quality results in response to many general and broad searches, thanks to its ESP search algorithm.
- ***Yahoo:*** Yahoo is the web's most popular search service and has a well-deserved reputation for helping people find information easily. The secret to Yahoo's success is human beings. It is the largest human-compiled guide to the web, employing 80 or more editors in an effort to categorize the web. Yahoo has at least 1 million sites listed.
- ***Google:*** Google (Figure 2) is a search engine that makes heavy use of link popularity as a primary way to rank web sites. This can be especially helpful in finding good sites in response to searches using generic terms such as 'vacation' or 'travel'.

All the search engines use the location and frequency of keywords in a

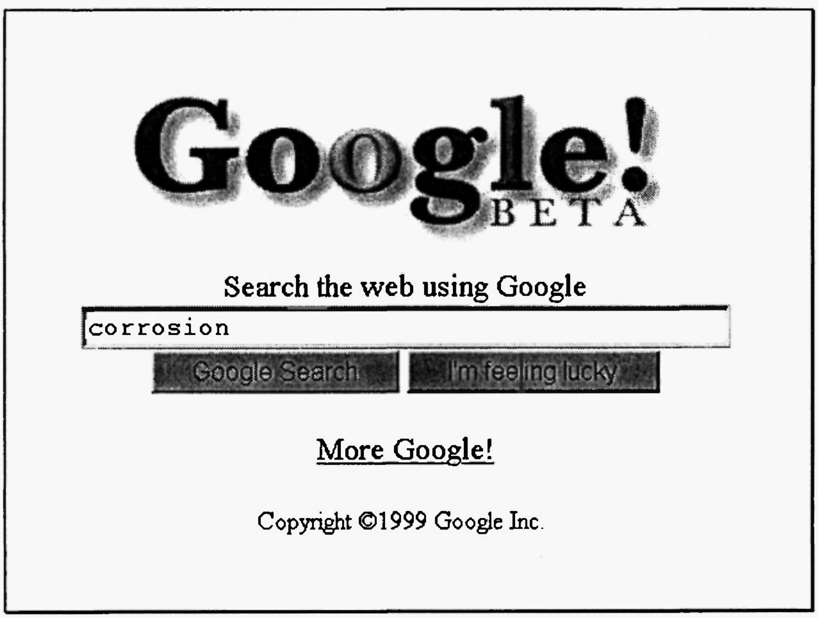

Fig. 2: Introductory page of a web search engine based on link popularity.

web page as the basis of ranking pages in response to a query. The exact mechanism is slightly different for each engine. In addition to location and frequency, some engines may give a page a relevancy boost based on other factors.

Well Traveled Corrosion Sites

The first ten most popular or traveled corrosion sites listed below have been identified with the help of the Google search engine. There is no apparent reason to explain the importance in traffic across these sites except that they obviously play some useful roles, not necessarily associated with a commercial slant. Many universities, for example, use the web as a billboard for every single activity related to University life, including the display of lecture notes and other studying material.

At least 7592 matches were found for corrosion with the Google search engine and that search took 0.78 seconds. The top ten corrosion sites are listed by decreasing order of popular linking:

1. The Electrochemical Society, Corrosion Division
   www.electrochem.org/divisions/corrosion.html
2. Corrosion Information Server Centre
   www.cp.umist.ac.uk/
3. Journal of Corrosion Science and Engineering
   www.cp.umist.ac.uk/JCSE/
4. Fontana Corrosion Center
   kcgl1.eng.ohio-state.edu/~frankel/fcc/
5. Smeltzer Corrosion Laboratory, McMaster University
   mse.eng.mcmaster.ca/resource/corrlab.htm
6. Intercorr Home Page
   www.intercorr.com/
7. Corrosion Research Center University of Southwestern Louisiana
   engr.usl.edu/crc/
8. Corrosion Engineering Consulting
   www.consultingengineers.com/Corrosion.html

9. Farwest Corrosion Control Company
   www.farwst.com/
10. Centre for Corrosion Technology
    www.shef.ac.uk/uni/projects/scr/

DEFINING INTELLIGENCE

Intelligence is one of those elusive concepts that are extremely context dependent. In a broad sense it is generally admitted that intelligence reflects the ability of an entity to learn, adapt and understand from experience. In that broad sense intelligence has been attributed to other biological systems than human beings (Stonier, 1992). A more human centered definition of intelligence would surely emphasize the understanding element of such a broad definition. The high level of adaptability required in the scientific discovery process would be a good example of higher intelligence.

Properties of Intelligent Systems

As a practical compromise for discussing machine intelligence one could describe intelligence as the capability of a system to achieve a goal or sustain desired behavior under conditions of uncertainty (Rzevski, 1995). By analogy with processes associated with human learning experiences, computer based intelligence has been classified into five key information processing capabilities (Goonatilake *et al.*, 1998). A general description of these capabilities can serve as a comparison basis for evaluating a system degree or type of intelligence.

- ***Knowledge acquisition:*** Knowledge acquisition is a process that involves eliciting, interpreting and representing the domain targeted knowledge. Experts usually acquire expertise through years of working on similar problems without necessarily being aware of the depth of their expert knowledge. Some AI methods are much better designed than others for reproducing the learning process as can be seen in Table II that contains a property ranking of five popular AI techniques (Goonatilake *et al.*, 1998).

Table II

Property assessment of different intelligent methods.

	Knowledge acquisition	Brittleness	High-level reasoning	Minimal reasoning	Explanation
Expert Systems	★	★	★★★★★	★	★★★★★
Rule Induction	★★★★	★★	★★★	★★	★★★
Fuzzy Systems	★	★★★★★	★★★	★★★★★	★★★★
Neural Networks	★★★★★	★★★★★	★	★★★★★	★
Genetic Algorithms	★★★★★	★★★	★★★	★★★	★★★

- ***Brittleness:*** An operational definition of brittleness would be the inability of an intelligent system to cope with inexact, incomplete or inconsistent knowledge. Brittleness could also be described as the opposite of the adaptability of a system to maintain a desired behavior even in the face of unforeseen changes in the system inputs. As shown in Table II, ESs are particularly poor performers in both this property and the previous one. In contrast, reasoning in neural networks involves the numeric aggregation of representations over the entire network. In neural nets, pseudo knowledge is distributed amongst the multitude of nodes that make up the different layers of a network.

- ***High level reasoning:*** High level reasoning refers to the ability to perform cognitive tasks. Humans solve problems by combining facts with knowledge. The human reasoning process consists of taking facts about a specific problem and, using them in combination with a general understanding of the problem domain, to derive conclusions. Understanding how humans reason, how they work with information on a given problem can guide the design of a system for processing knowledge.

- ***Low level or minimal reasoning:*** Low level reasoning relates to the ability to recognize patterns. Neural networks excel in the pattern recognition process. The only reasoning present in neural networks comes from the data preprocessing stages that are often necessary to condition data sets. Similarly to neural networks, fuzzy logic systems are also good at low level signal processing tasks such as industrial control and robotic sensing.

- ***Explanation:*** The ability to provide users with explanation of the reasoning process is an important feature of intelligent systems. It is an important aspect of intelligent systems to be able to explain how they arrive at a particular conclusion. In knowledge based systems, this information is often obtained by evaluating the trace generated by the inference engine. Neural networks are very poor performers in this property.

INTELLIGENCE IN CORROSION

From the previous discussion on machine intelligence properties, and in the light of the ranking presented in Table II, it is fairly obvious that the hopes of finding a system with rounded intelligence that would be built on a single technology are basically minimal. The following section presents some of the AI systems that have developed in recent years in various fields of corrosion prevention and control.

Expert Systems

Since the mid 80s, the transfer of corrosion expertise into expert systems (ESs) was attempted in a multitude of projects. Most systems reported were focused on the prescription, diagnosis and training for corrosion prevention. Only a few systems have dealt with the monitoring and planning aspects of corrosion prevention and control. Most systems were developed for distribution on personal computers (PCs), a different strategy than the practice reported for the early days of ES development.

Broad acceptance of the computer in the workplace has also facilitated the introduction of new concepts and methods to manage corrosion information. Very focused ES have been integrated in large systems as controllers or decision support systems to prevent corrosion damage. The progress and recent developments of ES related to corrosion prevention and control have been reviewed elsewhere (Roberge, 1994; Roberge, 1997). The following list indicates the major areas for which some systems have been reported:

- Cathodic protection;
- Cooling waters;
- Diagnostics;
- Inhibitors;
- Materials selection;
- Petroleum industries;
- Reinforced concrete;
- Risk analysis.

Neural Networks

An artificial neural network (ANN) is a network of many very simple processors or neurons, each possibility having a small amount of local memory. The interaction of the neurons in the network is loosely based on the principles of neural science. The neurons are connected by unidirectional channels that carry numeric data based on the weights of connections. The neurons operate only on their local data and on the inputs they receive via the connections. Most neural networks have some sort of training-rule. The training algorithm adjusts the weights on the basis of presented patterns. In other words, neural networks "learn" from examples. When ANNs inputs and/or outputs contain evolved parameters, their computational precision and extrapolation ability significantly increases and can even outperform more traditional modeling techniques. Only a few applications of ANN to solving corrosion problems have been reported so far. Some of these systems are briefly described here:

- ***Predicting the SCC risk of stainless steels:*** The risk of encountering a stress corrosion cracking (SCC) situation was functionalized in terms of the main environment variables (Smets *et al.*, 1992). Case histories reflecting the influence of temperature, chloride concentration, and oxygen concentration have been analyzed by means of a back-propagation network. Three neural networks were developed. One was created to reveal the temperature and chloride concentration dependency (Figure 3) and another to expose the combined effect of oxygen and chloride content in the environment. The third ANN was trained to explore the combined effect of all three parameters. ANNs were found to outperform, during this project, traditional mathematical

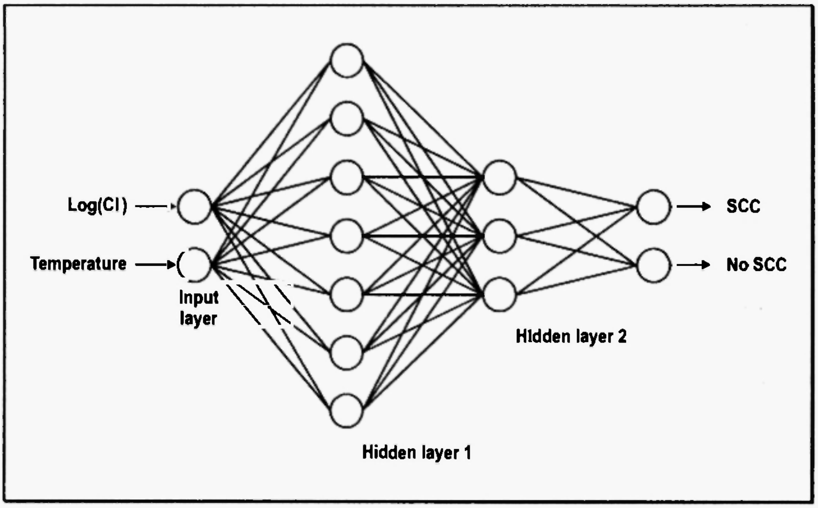

Fig. 3: Neural network architecture for the prediction of SCC risk of austenitic stainless steels exposed to industrial environments.

regression techniques where the functions have to be specified before performing the analysis.

- ***Corrosion prediction from polarization scans:*** ANN was put to the task of recognizing certain relationships in potentiodynamic polarization scans in order to predict the occurrence of general or localized corrosion such as pitting and crevice (Rosen *et al.*, 1992). The initial data inputs were derived by carefully examining a number of polarization scans for a number of systems and recording those features that were used for the predictions. The final ANN proved to be able to make appropriate predictions using scans outside the initial training set. The resulting ANN was imbedded in an ES to facilitate the input of data and the interpretation of the numerical output of the ANN.
- ***Modeling CO_2 corrosion:*** A CO_2 corrosion "worst case" model based on an ANN approach was developed and the model validated against a large experimental database (Nesic *et al.*, 1998). An experimental database was used to train and test the ANN. It consisted initially of six elemental descriptors (temperature, partial CO_2 pressure, ferrous and bicarbonate ion concentrations, pH, and flow velocity) and one output, i.e. the corrosion rate. The system demonstrated superior interpolation

performance compared to two other well known semi-empirical models. The ANN model also demonstrated extrapolation capabilities comparable to a purely mechanistic electrochemical CO_2 corrosion model.

- ***Predicting the degradation of nonmetallic lining materials:*** An ANN was trained to recognize the pattern between results from a sequential immersion test for nonmetallic materials and behavior of the same material in field applications (Silverman, 1994). Eighty-nine cases were used for the supervised training of the network. Another seventeen cases were held back for testing of the trained network. An effort was made to ensure that both sets had experimental data taken from the same test but using different samples. Appropriate choice of features enabled the ANN to mimic the expert with reasonable accuracy. The successful development of this ANN was another indication that ANNs could seriously aid in projecting laboratory results into field predictions.
- ***Validation and extrapolation of electrochemical impedance data:*** The ANN developed in this project had three independent input vectors, frequency, pH, and the applied potential (Urquidi-Macdonald *et al.*, 1997). The ANN was designed to learn from the invisible or hidden information at high and low frequencies and to predict in a lower frequency range than that used for training. Eight sets of impedance data acquired on nickel electrodes in phosphate solutions were used for this project. Five sets were used for training the ANN and three for its testing. ANN proved to be a powerful technique for generating diagnostics in these conditions.

Case Based Reasoning

Much of human reasoning is case-based rather than rule-based. When people solve problems, they frequently are reminded of previous problems they have faced. For many years, both law and business schools have used cases as the foundation for knowledge in their respective disciplines. Within AI, when one talks of learning, it usually means the learning of generalizations, either through inductive or through explanation-based methods. Case Based Reasoning (CBR) is unique in making the learning little more than a byproduct of reasoning (Kolodner, 1993). CBR has met with tangible success in such diverse human decision-making applications as banking, autoclave loading, tactical decision-making, and foreign trade negotiations. The CBR approach is particularly valuable in cases containing ill-structured problems, uncertainty, ambiguity, and missing data. Dynamic environments can also be tackled, or when there are shifting, ill-defined and competing objectives. A critical issue for the successful development of case based reasoners is the creation of a robust indexing. Three issues are particularly important in deciding on the indices (Barletta *et al.*, 1988):

- Indices must be truly relevant;
- Indices must be generalized, otherwise only an exact match will be the criterion for case applicability;
- But indices shall not be over-generalized.

Failure analysts and corrosion engineers also reason by analogy when faced with new situations or problems. Two CBR systems have been recently developed in support of corrosion engineering decisions. Both systems derived their reasoning from a combination of two industrial alloy performance databases. The first, "M-BASE", facilitates the process of retrieving materials according to a given set of desired properties and/or specifications. The second, "C-BASE", helps the materials engineer in the difficult task of selecting materials for corrosion resistance in complex chemical environments (Bogaerts *et al.*, 1997).

SUMMARY

The construction of intelligent systems can be quite complex when the domain of expertise is itself either complex or poorly supported by documentation. Many corrosion tasks fall into this category. A model of corrosion information for system life assessment would be incomplete if it did not include some of the multidisciplinary elements typical of a corrosion situation. The development of virtual expertise for fields related to failure analysis and a global framework that offers modularity and flexibility can guide maintenance engineering. When and where probabilistic models are developed, intelligence fusion techniques can help to design a logical representation of the materials elements controlling the lifetime of systems. In the meantime, it can provide a methodology by which information can be transferred across fields of expertise. Any progress in this direction of preserving information is enough motivation to pursue the development of better or more efficient models of expertise based on global information integration strategies. Information and intelligence are, after all, the most precious human commodity.

REFERENCES

Anonymous, *Beyond Information*. London, UK: Springer-Verlag. (1992).

Explanation-Based Indexing of Cases. Cambridge, MA: MIT Press. (1988) Proceedings of AAAI-88, 50-60.

Bogaerts, W. F., Smets, H., C.Sturrock, Arents, H. C., Zheng, J. H., and Barbaro, P. Corrosion risk assessment and computer-aided failure analysis. http://www.mtm.kuleuven.ac.be/Reports/ScientificReport92-94/Chapter8.html#Sec2. 1997.

Dreyfus, H.L. and Dreyfus, S.E. (1986) *The Power of Human Intuition and Expertise in the Era of Computers*. The Free Press, New York, NY.

Durkin, J. (1994) *Expert Systems: Design and Development*. MacMillan Pub. Co., New York, NY.

Goonatilake, S. and Khebbal, S. (1998) Intelligent Hybrid Systems, in: (Goonatilake, S. & Khebbal, S., eds.), *Intelligent Hybrid Systems: Issues, Classification and Future Directions*. John Wiley & Sons, New York, NY.

Kolodner, J. (1993) *Case-Based Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA.

Nesic, S. and Vrhovac, M. (1998) *Journal of Corrosion Science and Engineering*, **1**, 1-11.

Roberge, P.R. (1994) *Materials Performance* **33** , 52-56.

Roberge, P.R. (1997) *Corrosion Reviews* **15**, 1-14.

Rosen, E.M. and Silverman, D.C. (1992) *Corrosion* **48**, 734-745.

Rzevski, G. (1995) Applications of Artificial Intelligence in Engineering X, in: (Rzevski, G., Adey, R.A. and Tosso, C., eds.), *Artificial Intelligence in Engineering: Past, Present and Future*. pp. 1-16.

Silverman, D.C. (1994) *Corrosion* **50**, 411-418.

Smets, H.M.G. and Bogaerts, W.F.L. (1992) *Corrosion* **48**, 618-623.

Sullivan, D. *How Search Engines Work.* http://www.searchenginewatch.com/webmasters/work.html. 1998. (GENERIC) Ref Type: Electronic Citation.

Urquidi-Macdonald, M. and Egan, P.C. (1997) *Corrosion Reviews* **15**, 169-194.