KNOWLEDGE ENGINEERING IN PUBLISHING AND PRINTING

Aarne E.S. Klemetti *) and Simo T.P. Karttunen **)

Introduction

This paper describes how and why artificial intelligence (AI) will find many applications within the publishing and printing business. Knowledge engineering (KE) means the opportunity of using knowledge instead of plain information. In practical application this means acquiring knowledge from live experts or systems to form a **knowledge base**. A simplified model for the knowledge base is that it contains a skill base (heuristics and transforms) and a conceptual database (facts and structure) (see fig.1.). These can then be connected to the users by the modes of inference and the interpreter (objectives, goals, purpose) and then we have what is called a complete **expert system** (ES).

---

*) SISU Research, Inc.

**) VTT/Gra

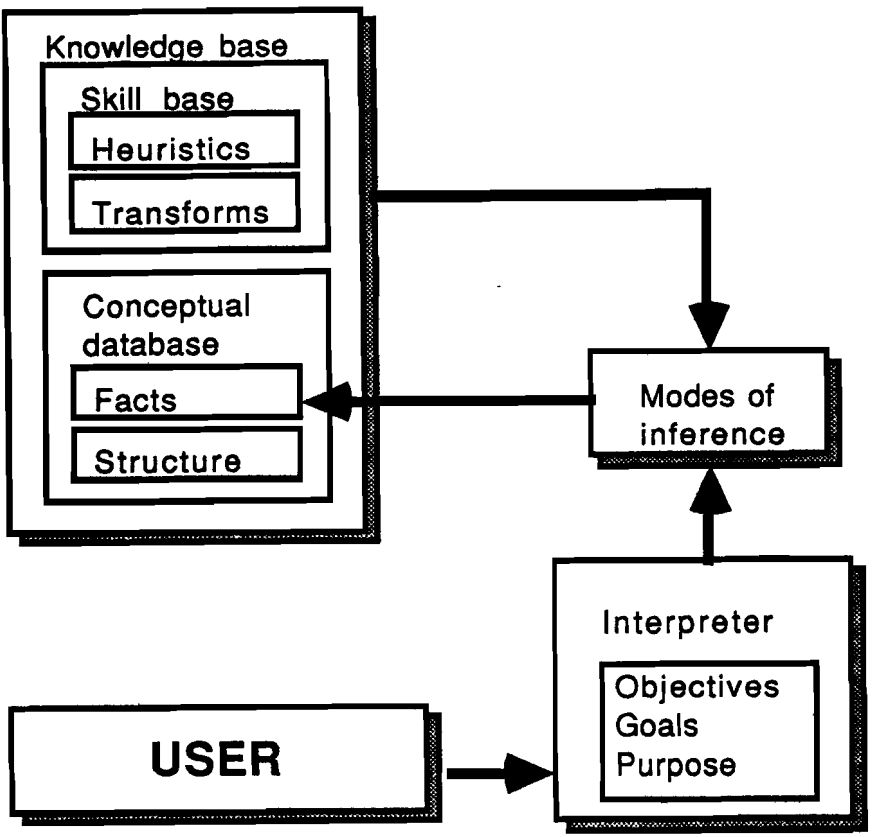

FIGURE 1. A conceptual model of knowledge base according to Addis (1985).

Basic concepts of artificial intelligence are presented in many handbooks of which only some latest ones are mentioned here (Addis, Feigenbaum, Harmon et al, Scown, Waterman, Winston). It is easy to get informed about the basic ideas of AI and KE. So far their applications are mostly experimental. The AI business is already well developed by many of the computer companies and specialized system vendors such as DEC, Inference Corp., IntelliCorp, Lisp Machine, Sperry, Symbo-

lics, TI and Xerox. The total sale of the AI products approaches USD 500 million this year, and it is estimated to quadruple before 1990.

There are as many definitions for the AI as there are prominent scientists in this wide field. The definition which is perhaps realistic enough is given by Patrick H. Winston:

> "The goals of the field of Artificial Intelligence can be defined as ...(attempting) to make computers more useful (and) to understand the principles that make intelligence possible" (Winston 1984).

The authors of this paper like this definition because it clearly emphasizes the real distance between human intelligence and our best machines.

Some AI tools are now commercially available. They are ready to march out of research groups and laboratories to real operational systems. We might also state that they will be used in acquiring, representing, accumulating and distributing expertise in most industries and business - expertise and knowledge which is vital but scarce in the ever increasing flood of unscreened - often useless - information.

Perhaps not yet directly applicable but famous and important are the so called **Fifth Generation Computer System** (FGCS) programs which were started by Japan about 5 years ago by forming the Institute for New Generation Computing (ICOT). Critical, adequate and updated reviews on this matter are in our references (Feigenbaum 1983) and (Scown 1985).

The understanding, meaning, parsing, syntax, translation and natural language problems in publishing can get some future solutions from these giant projects. Only the ever increasing need for commercial and technical translations requires the use of machines. The same can be said about the many user interface developments for editorial work in scientific, professional and daily publishing.

Many competitive and supportive international AI research programs such as MCC in USA (1982) Alvey in UK (1982) and Esprit in the European Community (1983) were started later which are dealing with Fifth Generation Computer and other central AI problems.

There has not been much discussion about the potential applications of AI or KE in **publication and printing** business. Some of our recent reports (Ahonen 1984, Karttunen et al, Karttunen, Klemetti 1985, 1986b) have mentioned AI as promising future technology. A review paper by Fuchs (1986) has been prepared for the IFRA members on the applications of AI in **newspaper production.** Ahonen and Mäki (1984) made some early attempts to define possible goals for AI and KE in publishing systems. From that time on our laboratory has made preparatory work and investments in publishing applications of AI for a start.

So far the only commercial products - or rather the first prototype systems - on the market are the Expert Publication System by CSI (CSI 1986, Fusco 1986) and Yellow Page Directory System by ETI (Gay 1986) which have been demonstrated in recent exhibitions. Several systems for various applications are **under development** some of which will be mentioned in the following chapters.

A new computer research centre ECRC announces to develop several applications of KE e.g. for interfacing relational databases to logical programming (PROLOG) and defining logically formulated request languages which can then be used to "handle deductive information, create dynamic overviews and promote dynamic synthesis of information" (ECRC 1986). ECRC is a joint basic research centre of three major computer companies in Europe the Bull (France), ICL (UK) and Siemens (FRG) located in Munich, developing new basic technologies to be later applied by the industries in their system products.

This paper is a brief survey of the current state of the first attempts to apply AI in publishing and printing business. We are also going to predict some most probable future goals and trends in these applications. We expect that the AI will be having a major impact on the publishing and printing systems and products during the early 1990's.

This paper also gives information about the progress of our own work in developing methods based on knowledge engineering (Klemetti 1986a), and related topics (Karttunen et al, Karttunen, Lehtonen, Visa). It is presumable that it will encourage other research and development groups in universities, research institutes and companies (of publication and printing sector) to use AI tools in their daily work.

On the existing approaches

So far there has not been too noticeable interest in the ES's in the graphic arts industry. However, some applications have seen the light. One of them - a typical ES - is the CSI Expert Publishing System, which is expected to solve the whole newspaper pagination problem. They have integrated conventional text, image and layout processing tools under the AI umbrella developed with Inference Corporation's ART (Advanced Reasoning Tool) (Art 1985). The whole system is connected by relational database management - DEC Rdb. That a rather natural approach with current generation software tools.

The CSI system itself is an ambitious attempt to use the decision support technology in the day-to-day operative production. The former DSS's were made to support executives in their what-if calculations. The whole idea was to give the user as many facts - axioms or conclusions - as possible. These facts or derivations should help the user to make optimal decisions. In CSI's case these facts include press configurations and other explicit things that are necessary people

who sell ads, for example. They have to know which pages can take color. The system can also suggest the newsroom user possible layouts for newshole filling. The CSI system does not solve implicitly or explicitly the fuzzy nature (Klemetti 1985) of page elements' - news' and ads' - internal shape. User is not able to get aesthetic evaluations on the items he/she is producing.

Problem oriented applications can also be found from image system applications. These devices, e.g. the Scitex Smart Scanner (Scitex 1986), can analyze the material to be input or output. They apply techniques called image analysis and pattern recognition. The purpose of these operations is to enhance the product: i.e. its shape, grey levels, contours, screens.

In this TAGA-conference a new plate imposition system will be presented by MGD Rockwell (Balban 1987). This is a natural part of computer integrated manufacturing in printing plants.

Editorial systems

In the editorial offices of newspapers or magazines there are many **language related problems** such as search, archiving, spell checking, and text-critiquing. Intelligent text retrieval systems are based on AI tools (ECRC, Rauch-Hindin, Schank et al, Sparck-Jones, Zarri) by using the morphology, syntax and semantics of the natural language.

Even human speech taken in by microphone can be translated with relatively large - i.e. around 1000 words - vocabularies (Scown 1985). **Speech recognition** cannot be used as an editorial input method for many reasons besides the lexical limitations. In the newsroom it is impractical to use speech to pass text into computer. However, limited applications in search or command applications are possible.

Natural language **understanding** in the case of large text databases of editorial background or archived information could increase intelligence far beyond the relational features. The research in the natural language (Rauch-Hindin, Scown, Schank et al, Sparck-Jones) will end up in many applications. The **machine translation** between the large languages such as English, French, Japanese, Russian and Spanish will be needed in the foreign trade, culture, education and communication (Slocum, Tucker). The international encyclopedia and dictionary publishing alone is a large scale target for editorial language systems. For example the Groliers encyclopedia is available on CD-ROM but there is no proper software for using it efficiently.

Most of the current **editorial input** systems are rough and not too sophisticated. Let us first look at the text input and processing systems. They have device-to-device connections but all the existing systems are passive and dumb. It is not possible to scroll around the databases of news services instead of just taking in all they send. This includes also the inhouse archives. People are confused by irrelevant questions like "how many gigabytes of memory is needed to store the material of a mid-sized newspaper" or "we can have the Grolier's in one CD-ROM, but when are we able to update it with software?"

What we need is dynamic and **intelligent** data management and retrieval systems. In editorial work it would be very useful to be able to gather data intelligently in the background. This is not very utopistic when we compare the whole process to the spelling check. The checker reads the text file and compares the words with existing dictionaries. If there are ambiguities, it flags to the user.

A new thing to be solved by applications of AI is the **semantic analyser**, which serves as a dynamic interface to databases. We can imagine it as a "customized hypertext" connection. The application should be able to read material, parse it and act with or without the user's

impact. The user interference to the keyword indication could be carried out for example by activating some words, sentences or graphics. Briefly this means that the user can link different input sources, like video, databases, news services and text and image archives, to his/her work.

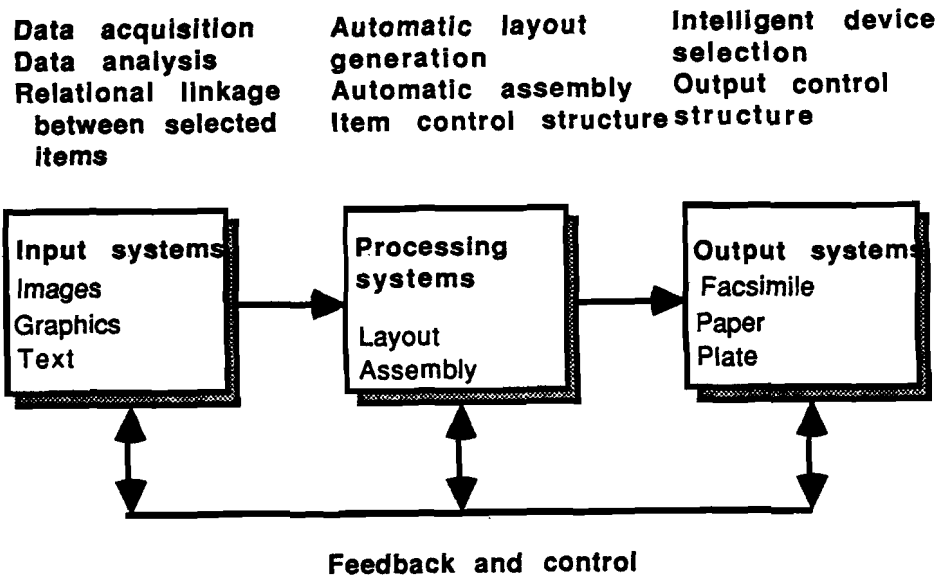

FIGURE 2. A hypothetical description of graphic arts production that incorporates the computer based expertise.

There is a strong demand on automating the layout process itself. Why not use AI to the assembly? Although the page assembly - regardless of the end product - is a heavy process in the conventional production, it is possible to be skipped over with automatic or semi-automatic design.

How to reach this goal? To understand the nature of pagination we must see that the final assembly can be embedded into the design process. If we think about it as defining operational instructions, we notice immediately that the instructions or definitions should be manipulated automatically.

For example, if we order the system to put item A into place (X1,Y1;X2,Y2) (co-ordinate definition for a square element), it should be placed there without any further instructions. The ultimate creative responsibility should be kept with the user. Thus the ES could inform and guide the user in the design phase, because there is nothing qualitatively complex in the assembly. There are just quantitative orders on how to place something somewhere. We should also be able to monitor the production. An conclusion from this is that the more integrated the whole pagination process is the more it employs a proper production management system - and KE.

Now let us look at the production control and management problematics. Again, we can divide it into subproblems, that can further be analysed quantitatively and qualitatively. For instance, the gathering of data is clearly a quantitative process, while the management of it is qualitative. The systems should always be able to interfere with the production when problems occur.

Ad marketing and production

In newspaper production advertising income is the economic backbone of the business. The ad marketing and production is a two-shifts-on-seven-days/week - i.e. a very sensitive - operation which in many larger newspapers fills more than half of the column space of the newspaper editions.

The ad marketing methods vary from one newspaper to another ranging from direct sales, to telephone ad-intaking and ad office services. There may be local service companies such as ad agents, typesetters or repro shops providing ready display ads, text, graphics or color separations.

The expertise in selling various classes of ads such as cars, flats, houses and real estates, clothes, consumer goods and vacant jobs, may consist of:

* Classified structures for better reader service
* Booking and selling routines in ad-intake
* Databases, tables or indexes for various ads
* Contact information of the frequent advertisers
* Telematic connections from the sites of advertisers

It is self-evident that each newspaper is trying to develop these most strategic marketing tools. At the same time the production of ads has been changing from plain ad composition towards more integrated systems including booking, billing, credit checking, material and production tracking and automatic page make-up.

The next step will be to develop expert systems for ad marketing and production control. This will be done in large newspapers having an adequate economy - i.e. dozens of daily ad pages - and possibly in joint pilot projects with some systems vendors.

The expert system could easily optimize the use of sales resources and maybe improve the service level to the advertisers. The AI systems could adaptively learn so that the frames and production rules could be dynamic according to measured ad effects on final sales of the advertisers. Similar systems have been used in financial services as one of the first commercial applications of AI (Scown 1985).

When the advertising business moves towards the more targeted new methods of direct marketing newspapers and magazines get a new challenge. This may be more diversified or even personalized publishing products with the corresponding ad services. However, such systems need much more information and intelligence in order to be profitable.

An AI based ad marketing system might consist of a knowledge base including facts about the customer groups, demographics, competitors, market geography, annual events, important seasons and statistics. These could be in a form of relational databases and used by the rule and frame based KE tools. Such a system might also lead to more targeted editorial products such as personal or on-demand newspapers which may be a future trend or a byproduct (Burkhardt 1987).

Paper and print quality

Paper formation has been a term used for the visually observable flocculated structure which is apparent when viewing paper in transmitted light. The uneven formation depends on the degree and structuralities of the variation in mass of the fibrous and other material of paper. For quantitative measurement of formation the paper physicists have been using simple statistics such as standard deviation or mean deviation of the formation signal - spotwise measured transmittance or mass variation. Image analysis of a high resolution formation image using signal processing algorithms has been proposed by Visa (1986).

Paper formation is known to be a critical paper property for many print quality factors such as evenness of solid and halftone prints, back-trapping mottle and print through. Formation and formation-related properties can be measured by analyzing the surface or transmission image, beta-radiographs or various prints as reviewed by Bery (1985) who also developed a dry graphite tester and its print analyzer. Transmission formation test has also been included in the measurements of robotized paper laboratory of the RoboTest (PaperLab 1986). The measurement and its PaperLab-module, developed jointly by RoboTest Company and Graphic Arts Laboratory of VTT, is based on image analysis (Karttunen et al, Visa).

Print quality analog to paper formation can be measured with the CCD sensors as shown by the preliminary tests about three years ago by Launonen (1984). The speed and quality problems of the CCD sensors are the topics of a paper by Simomaa (1987) in this conference. Signal processing, digital filtering and image analysis methods and software can be used to compress the massive amount of signals (Karttunen et al, 1987). The intelligent choice of relevant measuring strategies during each print run sequence (start, run, disturbancies, stop) and print job (webbing, colors, page make-up, run length, quality level) becomes a natural target for AI methods in integrated print quality control.

Intelligent image measurement methods are under development and they may be expected (Karttunen et al., Lehtonen 1987) to substitute the existing measurements of print quality from test targets - by "spotwise" densitometers (on- or off-line). This is particularly necessary in newspaper printing where several webs are used and no test pattern can be printed. Even in low speed tests (Launonen 1984) the requirements for both sensor and signal electronics become very high. The sensors and particularly image processing and analysis electronics are getting more powerful each year and they give us prospect of new solutions.

Another approach to print quality studies by Saarelma and Oittinen (1986) uses new definitions of print quality based on systems analysis. These methods can be combined with the fast image capture and analysis. In the measurement of an integrated system environment (control system) new definitions may work even better than the traditional print quality parameters. We need new concepts and much more measuring power and intelligence in the quality control of originals, reproductions, proofs and prints (Karttunen et al, Lehtonen). Such new approaches are on their way and they are the only way to the total mill level quality management (Lehtikoski 1986).

Computer vision and image recognition are wide fields of research and applications (Fu et al. 1984). There will be many quality control and visual inspection applications of intelligent computer vision for paper and printing plants in the near future (Karttunen et al, Lehtonen, Lehtikoski, PaperLab). If the complicated images of microelectronic circuit diagrams can be understood, could then the print quality of any prints be impossible or too difficult ? In applications of image analysis we have already applied AI methods (Ahonen et al, 1985) and there are no particular limitations for developing the intelligence of print quality control systems.

Conclusions

The whole field of artificial intelligence and its applications is very exciting. We must understand that AI will never totally replace the procedural and algoritmic data processing. Its role is merely complementary. The more the tools develop the more they will be integrated. As we can see from the Japanese FGCS project, their first goal, which they have already accomplished, is the relational database management. They solved it by designing a relational database management computer.

One very difficult problem in ES development is the lack of proper design methods. We have

framework for developing and analyzing a "conventional" system and a database system, but only some more or less inconsistent guidelines for designing an ES. That is one reason why we have knowledge engineers and are - if possible - more dependent on their skills than we used to be on the "ordinary" computer experts.

What has driven us to this situation? In the beginning of this decade it was very popular to predict that before the end of the 80's 95 % of US. population will be employed by software code writing. The development tools have reached the stage that programmers are much more productive than they used to be a couple of years ago, thanks to the workstation architecture and graphic user interfaces.

There are several different developments for applying artificial intelligence and knowledge engineering tools in publishing and printing systems. So far only some of the first prototypes are available as commercial products. The language and other human-machine interface methods of AI are natural sources for new applications in publishing and printing because they both deal with speech, meaning, knowledge, recognition, audio/visual perception or information retrieval.

We have gathered available information about the first attempts made by our own groups and others. The applications will range from editorial and design through ad marketing and production systems to printing and mill level quality and production control. Many systems will be needed since they are the only way to solve really difficult and often heuristic problems. In some cases the need for more flexibility and speed will be a decisive motivation for using AI tools.

In addition to the editorial applications, which are many, the AI can be applied in automatic paper and print quality control. The production control and maybe also the preventive maintenance systems could be improved by adding intelligence to their software.

Man-hours are the most valuable resources used for publishing and printing. This labor intensive industry must try to improve its productivity. The degree of automation is fairly high now but new methods such as knowledge engineering can still improve the most difficult tasks including the creative professions.

Literature

Addis, T.R.
1985 Designing Knowledge-Based Systems. Kogan Page Ltd, London.322 pp.

Ahonen, H. and Mäki, S.
1984 Artificial Intelligence in the Production of Graphic Industry. STEP-84 Symposium Papers. Otaniemi, Espoo. p. 279-286.

Ahonen, H., Jacobson, S. and Ylinen, R.
1985 Proceedings of the 2nd Int. Conf. on Artif. Intelligence Dec 1-5, 1986. Marseille. p.609-621. Dr Ahonen is a member of the staff of VTT/GRA.

ART
1986 The ART System Product Information. Inference Corp.

Balban, M.
1987 The Determination of Printing Press Imposition by Means of Artificial Intelligence Technologies. TAGA 87 Conference. San Diego CA.

Bery, Y.,
1986 Graphite Tester. Conceptual Development. American Ink Maker vol. 64(1986) no.7, pp.28-37.

Burkhardt, F.
1987 A Personally Profiled Newspaper. Paper given in TFL 25th Anniversary Symposium Lidingö, Sweden. Jan 1987. To be

published in the symposium proceedings by the TFL Stockholm. Mr. Burkhardt is the president of IFRA.

CSI
1986 AI-based Expert Publication System. Product information by Crosfield CSI.

ECRC
1986 The European Computer-Industry Research Centre in Munich. Bull S.A. Insert no.02/86.

Feigenbaum, E.A. and McCorduck, P.
1983 The Fifth Generation. Reading, MA. Addison-Wesley.

Fuchs, B.
1986 AI Expert Systems and their Applications in Newspaper Production. IFRA Special Report no. 3.3 , Darmstadt, 26 pp. Mr. Fuchs is the research director of IFRA.

Fu, K.S. and Rosenfeld, A.
1984 Pattern Recognition and Computer Vision. IEEE Computer, Oct 1984, p. 274-282.

Fuchs, B. and Weide, W.
1986 Automatic Reel Transport by means of AGVS. IFRA Special Report no. 3.4 , Darmstadt, 29 pp.

Fusco, R.
1986 Artificial Intelligence for Newspaper Pagination and Control. Lasers in Graphics '86 Conference Proceedings vol. 2, Session 6-1C, Anaheim, CA. pp.349-351.

Gay, J.
1986 The Application of Expert Systems to Pagination. Lasers in Graphics '86 Conference Proceedings vol.2, Anaheim, CA. pp.352-356.

Haltsonen, S.
1986 Trends in Speech Recognition. STeP-86 Symposium Papers. Espoo.

Harmon, P. and King, D.
1985 Expert Systems, Artificial Intelligence in Business. John Wiley et Sons, New York.

Karttunen, S., Karttunen, A., Lindqvist, U., Södergård, C.
1987 Intelligent Image Analysis: Measuring Paper Formation and Print Quality. Graphic Arts in Finland vol.16 (1987) no.2. Ms A. Karttunen, Dr Lindqvist and Mr Södergård are members of the staff of VTT/GRA.

Karttunen, S.
1984 New Aspects on Electronic Publishing. IARIGAI Management Conf. May 1984, St Gallen, Switzerland. A preprint of 7 pp. Dr Karttunen is the director of the VTT/GRA and co-author of this paper. Docent at the Helsinki University of Technology.

KEE
1985 The KEE System Product Information. IntelliCorp.

Klemetti, A.E.S.
1985 Automatic Newspaper Pagination - A Dream or Reality? Lasers in Graphics '85. Nashville, TN. Klemetti is a co-author of this paper. He worked for VTT/GRA between 1983-87. From 1987 on he is research manager of SISU Research, Inc.
1986a. Expert System in Newspaper Pagination. STeP-86 Symposium Papers. Espoo 1986, 7pp.
1986b. A Vision of the Optimized Pagination Environment. Lasers in Graphics '86. Anaheim, CA.

Launonen, R. Kaivosoja, J. and Lehtonen, T.
1984 Online Monitoring of the Newspaper

Print Quality. TAGA 1984 Conference.

Lehtikoski, O.I. and Ingman, L.C.
1986 Improved Competitiveness through New Technology in Quality Management. TAPPI Journal vol.69 (1986) no. 11, p. 64-68. Lehtikoski is the founder and president of RoboTest Oy of Finland.

Lehtonen, T., Launonen, R.
1987 Personal discussions on unfinished project on print quality monitor based on new image module and image analysis. Espoo, 1987. To be reported in 1988.

Negoita, V.
1985 Expert Systems and Fuzzy Systems. The Benjamin Cummings Publishing Co., Inc. Menlo Park, CA.

Pearl, J.,
1984 Heuristics. Addison-Wesley.

Rauch-Hindin, W.
1984 Natural Language: An easy way to talk to computers. Systems & Software, Jan 1984, p.187-230.

PaperLab
1987 PaperLab product information. Robotest OY.

Saarelma, H. Oittinen, P.
1986 Definition of Print Quality. 4th Tappi/CPPA Conference. Oct 1986. Ottawa. Saarelma and Oittinen are professors of the Helsinki University of Technology.

Schank, R. et al.
1980 Conceptual Information Retrieval. Yale Univ. Dep. of Computer Sc. Res. Report no. 190, Dec 1980.

Scitex
1985 Smart Scanner product information. Scitex Ltd.

Scown, S.J.
1985 The Artificial Intelligence Exper. An Introduction. Digital Equipment Corporation, Maynard , Mass. pp. 183.

Simomaa, K.
1987 Are the CCD Sensors Good Enough for Print Quality Monitoring? TAGA 87 Conference, San Diego Ca, March 1987. Dr Simomaa is a member of the staff of the Instrument Laboratory of VTT.

Slocum, J.
1985 A Survey of Machine Translation: Its history, current status and Future Prospects. Computational Linguistics no.11, 1985, p.1-17.

Sparck-Jones, K.
1984 Natural Language Interfaces for Expert Systems. In M.A. Bramer (Ed.) Research an Development in Expert Systems. Cambridge University Press, Cambridge. 1984.

Tucker, A.B.Jr.
1984 A Perspective on Machine Translation: Theory and practice. Commun. of the ACM, Apr 1984, 27(4), p. 322-329.

Visa, A.
1986 The Characterization and Measurement of Paper Formation with Visible Light. Signal Proc. III: Theories and Appl. I.T. Young et al.(Ed.). Elsevier Science Publ. B.V. (North-Holland). EURASIP, 1986, p.635-37. Mr. Visa is a member of the VTT/GRA staff and is hired by the Pöyry International for a one year project (1986-Oct 1987).

Waterman, D.
1985 A Guide to Expert Systems. Addison-

Wesley, Reading, MA. 1985.

Winston, P.H.
1984 Artificial Intelligence (2nd ed.) Addison-Wesley. Reading, MA. 1984.

Zarri, G.P.
1984 Expert Systems and Information Retrieval: An experiment in the domain of biographical data management. Int. J. Man-Mach. Studies, Jan 1984, 20(1), p. 87-106.

Zimmermann, H.
1986 Fuzzy Set Theory - and Its Applications. Kluwer Nijhoff Publishing, Netherlands. 1986.

APPENDIX 1

On the ES activities in VTT/GRA

VTT has growing AI culture in the most prominent laboratories, including the graphic arts laboratory. At the moment there are 8 Symbolics's and a couple of Xerox AI Stations. There have been experiments on testing different software tools for both micro-computers and larger installations. However, the most used tool is Intellicorp's KEE (1986).

Our laboratory has been involved with video image analysis application. Our part of this problem was solved using T-Prolog in DEC VAX (Ahonen and Mäki 1984). Six months ago we purchased a Symbolics AI workstation and KEE software to help our system development. Our objective is to test and develop methods and prototypes for industrial applications.

Graphic arts laboratory is currently developing a system for page design. It is based on the idea that the phase of the work which requires most of the human inference - the layout creation - must be automatized. It is quite understandable that the sorting and other administrative operations should not be manipulated by this system. The reason for this is the good old "divide-and-conquer" principle. One should not mix the expert system with pure calculations.

We have excluded everything but the publication and layout design. What happens after the creative phase is outside the scope of this project. Anyway, the nature of the page layout is very complex. The system should be able to generate alternatives of the layout or assembly model.

At the moment we are running test material on the system. The first stage is to design an optimal assembly model for rigid material, i.e. yellow page directories.

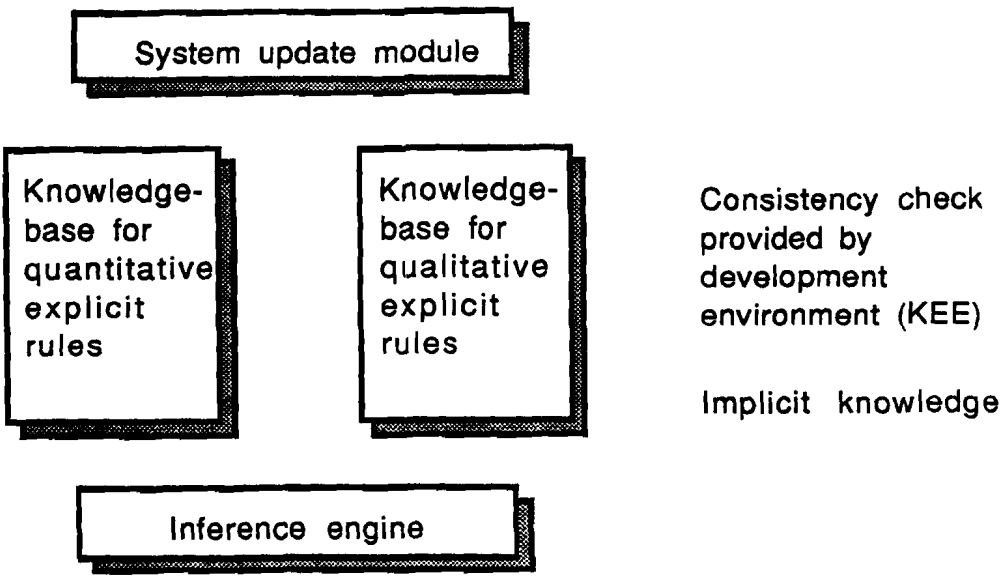

FIGURE 3. A schematic description of the page designing expert system, which is under development in VTT's graphic arts laboratory.