we write part-of(C,A)

E.g.: part-of(NERVOUS-SYSTEM, BODY)
part-of(HEAD, BODY)
part-of(TITLE, 2, BOOK)

element-of(E,G): each instance of G is a set of instances of E. E.g.:
element-of(STUDENT, CLASS)
element-of(PERSON, AUTHORS)

b) Predicates between instances

is-part-of(c,i,a): c is the i-th component of the aggregated entity a. E.g. is-part-of('Data Base Management', 2,(805367802):BOOK) 'Data Base Management' is the second component of the book identified by the ISBN below.

is-element-of(e,g): e is an element of the group g. E.g. is-element-of(('McFadden'),(805367802:ISBN):AUTHORS) McFadden is one of the authors of the group of authors identified by the booknumber below.

AXIOMS

The predicates below are characterized by the following axioms:

1. is-a(A,B) & is-a(B,A) <=> A=B
2. is-a(A,B) & is-a(B,C) => is-a(A,C) therefore is-a is a partial order relation
3. is-a(A,B) & is-a(A,C) => $\exists$ D(is-a(B,D) & is-a(C,D))
4. part-of(A,B) => ~ part-of(B,A)
5. part-of(A,B) & part-of(B,C) => part-of(A,C)
6. is-a(A,B) & part-of(B,C) => part-of(A,C)
   The case part-of(A,B) & is-a(B,C) => part-of(A,C) is not always true because not all elements of C must be aggregates with elements of A. But, if we consider a predicate **covering(r,C)** meaning that each element of C is in at least one subclass of C by role r, then we have the axiom
   6' ($\forall$ B(is-a(B,C,r) => part-of(A,B)) & covering(r,C) => part-of(A,C)
7. is-a(A,B) & part-of(A,C) => part-of(B,C)
8. is-a(B,C) & part-of(A,C) => part-of(A,B)
9. is-a(A,B) & element-of(B,C) => element-of(A,C)
10. is-a(B,C) &element-of(A,C) => element-of(A,B)
11. instance-of(e,C) & is-a(C,D) => instance-of(d,D)
12. is-part-of(c,d) & instance-of(c,C) & instance-of(d,D) => part-of(C,D)
13. part-of(C,D) & instance-of(d,D) => $\exists$ c(instance-of(c,C) & is-part-of(c,d))
14. (is-element-of(e,g)=>instance-of(e,E)) & instance-of(g,G) => element-of(E,G)
15. instance-of(g,G) & element-of(C,G) => ($\forall$ e(is-element-of(e,g) => instance-of(e,C))

As the IS-A relationship is a partial order we speak about subconcepts (or subclasses) and (super-)concepts. If is-a(A,B) then A is a specialization of B and each instance of A inherits all its own properties as instances of B. This means that the specialization is used to add new properties to existing ones. The inheritance of properties and property values can be given by two additional axioms:

a) Inheritance of properties (or attributes or relationships). If C(p) means that each instance of C has property p (e.g. PERSON(age), ANIMAL(genitor)) then

14) C(p) & is-a(B,C) => B(p)

b) Inheritance of property values. If C(p,x,y) means that instance x of C has value y under property p then

15) C(p,x,y) & is-a(B,C) => B(p,x,y)

The axioms presented here states a clear semantics of the hierarchical connections between concepts or entity classes. This semantics is fundamental for a correct modeling of the real workd and the axioms can be used for query processing in knowledge bases using semantic networks.

REFERENCES

[Br-83] R.J. Brachman, "What IS-A and Isn't: An Analysis of Taxonomic Links in Semantic Networks", IEEE Computer, October 1983.

[Ga-88] P.K. Garg, "Abstraction Mechanisms in Hypertext", CACM Vol. 31 no. 7, 1988.

[HM-78] M. Hammer & D. McLeod, "The Semantic Data Model: A Modelling Mechanism for Data Base Applications", Proc. SIGMOD Intl.Conf. on Management of Data, 1978

[Sc-83] U. Schiel, "An Abstract Introduction to the Temporal-Heirarchic Model (THM)", Proc. 9th VLDB, Florence, 1983

[SS-77] J.M. Smith & D. Smith, "Database Abstractions: Aggregation and Generalization" ACM TODS Vol. 2, no. 2 1977.

AI Online: The Coverage of the Literature of Artificial Intelligence in Online Bibliographic Databases

William J. Mills
Library and Information Services
Aston University, Aston Triangle
Birmingham, UK

Two years ago I analysed the degree of scatter of AI literature across a wide range of online bibliographic databases [1]. This paper presents the results of a similar survey two years later. Besides drawing attention to those databases which are most likely to yield relevant information for any particular query, the survey also demonstrates the massive growth in the amount of AI-related literature and to some extent identifies where that growth has been most rapid and which other areas may be ripe for development.

As in the previous study, the method used relies on the provision of cross-file index search facilities available on certain online host systems, and in particular on the Questindex, Dialindex and CROS facilities of ESA- IRS, Dialog and Data-Star host systems respectively. In all 58 databases were searched against 53 in the earlier study. To ensure that the two surveys were fully comparable exactly the same search statement was entered: ARTIFICIAL INTELLIGENCE OR EXPERT SYSTEM/S.

Table 1: Database Coverage of the AI Literature

	1986		1988	
Database	No.	Rank	No.	Rank
Inspec	5073	1	13810	1
Compendex Plus	3171	2	9057	2
Computer Database	2267	4	4180	3
NTIS	2413	3	4010	4
NASA	2058	5	3679	5
Prompt			3599	6
Pascal	1914	7	3223	7
Mathsci	1936	6	2872	8

Information Science	835	8	2805	9
Japan Technology			2074	10
ABI/Inform	531	10	1836	11
Scisearch	769	9	1557	12
Artificial Intelligence	379	15	1325	13
Microcomputer Index	403	13	790	14
ERIC	459	11	789	15
Social Scisearch	421	12	733	16
Robomatix	398	14	647	17
Dissertations Abstracts	121	19	529	18
Biosis	116	20	491	19
Psychinfo	199	16	487	20
Ismec	186	17	463	21
Embase	123	18	375	22
CAD/CAM	81	25	343	23
Medline			321	24
Lisa	113	21	319	25
Management Contents	84	24	271	26
Chemabs	112	22	260	27
Language Abstracts	35	27	225	28
Intime f			222	29
Philosophers Index	102	23	221	30
CEA	27	30	121	31
Agricola	25	31	118	32
Geobase			109	33
Georef	44	26	107	34
Ibsedex			98	35
Metadex	8	36	92	36
Brix	0	46	88	37
Sociological Abstracts	18	33	81	38
IRRD	10	35	72	39
Legal Resource Index	25	31	67	40
CAB Abstracts	7	38	64	41
Fluidex	11	34	57	42
PAIS International	28	29	51	43
Telecommunications Abstracts	8	36	45	44
Soviet Science & Technology	29	28	34	45
Analytical Abstracts			31	46
Enviroline	5	39	31	47
Molars	0	46	30	48
Geoarchive	2	42	26	49
Oceanic Abstracts	3	40	23	50
Pollution Abstracts	2	42	17	51
Energyline	0	46	14	52
Aluminum	0	46	13	53
Food Science	0	46	7	54
Int. Pharmaceutical Abs.	1	44	7	54
Meteorolog. & GeoAstro. Abs.	1	44	7	54
Delft Hydro	3	40	5	57
BNF Metals	0	46	1	58

Immediately apparent is the size of the increase in the amount of AI literature, which has more than doubled in the 2 years on many databases and has increased at substantially faster rates on several others. In all, there were 62,810 hits in the June '88 survey as against 24,560 in June'86. Of the 62,810, 6,345 were identified on the 7 databases not previously searched, but even with these subtracted, the overall new total is 56,475. For reasons to be outlined below, too much significance should not be attached to these exact numbers but the scale of the change is significant. Is there any other field which could currently manifest a similar rate of expansion?

Reinforcing this point, it is also noticeable that whereas 13 databases in the '86 survey showed less than 5 relevant items, in '88 only 1 database contained less than 5, and none held no relevant literature at all as against 6 in '86. Not only is the literature of AI expanding at a prodigious rate, it is clearly also expanding territorially and colonizing new areas of activity.

53 databases were included in the '86 survey. All were analysed again with the exception of Agris which was no longer available for a cross-index search. In addition, EI Meetings (ranked 5th in the '86 survey) and Compendex Plus. 7 new databases were added in the '88 survey. These are: Prompt, Japan Technology, Medline, Intime, Geobase, Ibsedex and Analytical Abstracts. Two of these contained substantial amounts of AI material: Prompt and Japan Technology.

Prompt performed surprisingly well considering that a study of its sources showed coverage of no core AI literature. Despite this, 3599 relevant items were identified making it the 6th ranking AI database. Clearly Prompt should not be ignored when searching for AI literature especially since it is likely to pick up aspects overlooked by other databases such as marketing and product information. Japan Technology contained 2074 hits. Given that it is only searchable from 1986, this is clearly an important database and likely to become ever more so, which is not surprising considering the strength of the Japanese AI effort.

Apart from these newly added databases, the same databases that performed best in '86 in general continue to do so. This is true particularly of Inspec and Compendex Plus which if anything appear to have increased their dominance.

For reasons, given in the '86 study, the survey's method inevitably under states the number of relevant items in any database, and overstates the total amount of literature across all the databases. The former because in any given database it is likely that there will be much AI-related material not using the search terms 'artificial intelligence' or 'expert system', which will therefore not be retrieved by the crude cross-index search. (This is likely to be especially true for much core AI material in core AI databases.) The latter because the same items may occur in more than 1 database. The exact extent to which such duplication occurs is difficult to establish.

The '86 analysis allowed for a database duplication factor of 2.456 and estimated a total AI literature of 10,000 items. There is nothing scientific about this duplification factor and the reasons for adopting it are outlined in my earlier paper. Applying it to the '88 figures results in an adjusted estimated total AI literature of approximately 25,500. In Table 2 below, to facilitate comparison between '86 and '88 figures, I have eliminated the contribution of the 7 new databases and hence use an estimated total of 23,000. All these figures should be taken in the truly approximate spirit in which they are offered but the exercise produces instructive results.

Table 2: Adjusted Database Scatter

Database	'86 % of 10,000	'88 % of 23,000
Inspec	50.7	60.0
Compendex Plus	31.7	39.4
Computer Database	22.7	18.2
NTIS	24.1	17.4
NASA	20.6	16.0
Pascal	19.1	14.0
Mathsci	19.4	12.5
Information Science	8.4	12.2
ABI/Inform	5.3	8.0
Scisearch	7.7	6.8

Artificial Intelligence	3.8	5.8
Microcomputer Index	4.0	3.4
ERIC	4.6	3.3
Socscisearch	4.2	3.2
Robomatix	4.0	2.8
Dissertations Abstracts	1.2	2.3
Biosis	1.2	2.1
Psychinfo	2.0	2.1
Ismec	1.9	2.0
Embase	1.2	1.6
CAD/CAM	0.8	1.5
Lisa	1.1	1.4
Management Contents	0.8	1.2
Chemabs	1.1	1.1

Table 2 identifies 24 databases with 1% or more of the AI total. The databases listed range from the specifically AI, through primarily computer science (Inspec and Computer Database), to special subject databases such as ABI/Inform (business and management), ERIC (education), Biosis (biology), Ismec (mechanical engineering), etc. For a full subject breakdown see Table 3.

As noted above, Inspec and Compendex Plus remain significantly the most important databases, with Inspec in particular outstanding. Table 2 suggests that their degree of dominance if anything is increasing, which is good news for the subject searcher, since clearly the better the coverage provided by a few sources the greater the ease of searching and the higher the probability of significant material not being overlooked. There is likely to be much duplication between Inspec and Compendex Plus so searching the two together will unfortunately not yield the 99.4% suggested by Table 2 on the basis of no duplication. Rather, if only two databases can be searched, almost certainly the best options will be to search Inspec together with NTIS, NTIS being strong where Inspec is weak - namely in its coverage of the report literature. There is seldom much duplication between Inspec and NTIS, so a search of the two could well retrieve approximately 75% of the relevant literature.

After Inspec and Compendex Plus, the next 5 databases all show shrinking percentages of the total. This decline in proportionate coverage may partly reflect the disproportionate growth of the top two databases, but it is interesting to consider other reasons why they may not have expanded at equivalent rates. NTIS and NASA primarily consist of reports of US Government-funded research. Are there fewer such projects now being conducted, or is it that fewer are publicly listed because of their greater military and commercial sensitivity? Pascal provides good coverage of continental European work. Does Pascal's comparatively slow rate of AI expansion reveal a slower pace of AI development there? Mathsci's slower growth probably results from AI's progress as an applied and no longer purely theoretical study. In my previous paper I predicted that the Computer Database might soon become a genuine competitor to Inspec as the major AI source. This clearly has not occurred and is something of a disappointment.

The survey also provides instructive evidence indicating into which areas AI applications are spreading most rapidly, and those which may - possibly - be ripe for development. Table 3 classifies the databases into broad subject groupings. Not surprisingly engineering applications continue to grow most rapidly, being picked up in the literature of Inspec and Compendex Plus in particular but also in special subject databases such as Robomatix, Ismec, CAD/CAM, etc. Business and management also shows great growth with Prompt, ABI/Inform and Management Contents. AI applications seem to be taking off in building and construction. In '86, BRIX showed no AI literature, this year's survey found 88. Considering the early success of this year's survey found 88. Considering the early success of PROSPECTOR, AI interest in the earth sciences has lagged. Similarly, pharmacy has been slow to develop AI applications when compared with other areas within the life and medical sciences. Chemistry is another possible under-performer.

Table 3: Databases Grouped into Subject Categories ('88 rankings shown in brackets)

GENERAL	
Inspec (1)	NTIS (4)
Pascal (7)	Japan Technology (10)
Scisearch (12)	Dissertations Abstracts (18)
Soviet Science&Technology(44)	
ENGINEERING	
Inspec (1)	Compendex Plus (2)
Robomatix (17)	Ismec (21)
CAD/CAM (23)	Intime (29)
CEA (31)	Metadex (34)
Fluidex (41)	Telecommunications Abs. (43)
BUSINESS&MANAGEMENT	
Prompt (6)	ABI/Inform (11)
Management Contents (26)	
MEDICAL & LIFE SCIENCES	
Biosis (19)	Psychinfo (20)
Embase (22)	Medline (24)
CAB Abstracts (40)	Enviroline (46)
Int. Pharmaceutical Abs. (53)	
EARTH SCIENCES	
Geobase (33)	Georef (34)
Molars (47)	Geoarchive (48)
Meteorolog&GeoAstroAbs(53)	Delft Hydro (56)
COMPUTER SCIENCE	
Inspec (1)	Computer Database (3)
Artificial Intelligence (13)	Microcomputer Index (14)
MATHEMATICS	Mathsci (8)
LIBRARY & INFO. SCIENCE	
Information Science (9)	Lisa (25)
EDUCATION	Eric (15)
SOCIAL SCIENCES	
Social Scisearch (16)	Psychinfo (20)
Sociological Abstracts (37)	Legal Resource Index (39)
PAIS International (42)	
HUMANITIES	
Language Abstracts (28)	Philosophers Index (30)
CHEMISTRY	
Chemabs (27)	CEA (31)
Analytical Abstracts (45)	
BUILDING & CONSTRUCTION	
Ibsedex (35)	Brix (37)

To do a comprehensive AI search using the minimum number of databases, I would suggest searching Inspec, to retrieve the bulk of the academic journal and conference paper literature, NTIS for reports, and Dissertations Abstracts for theses. (Note Dissertations Abstracts' near doubling of its AI percentage. AI is manifestly a 'hot' PHD subject!). Prompt should be added if product and marketing information are required.

In addition, where searching for literature relating to specific applications, the special subject databases, such as ABI/Inform, Chemabs, Biosis, etc. should also be searched.

Such a strategy would not guarantee the retrieval of 100% of the literature but it is unlikely that anything of substance would be missed. Given constraints of finance and time, a search of Inspec, NTIS and the most relevant subject database for an applications search, should retrieve the vast majority of what it is truly necessary to see and will be sufficient for most purposes.

This pattern could well change substantially in the near future when the Turing Institute database becomes publicly available. Almost certainly the world's largest specialist AI database, this consists of 45,000 records (Sept. '88), most but not all AI-related, and is growing at a rate of 1,200- 1,500 per month. (These figures should be compared with my estimate of 25,500 for the total AI literature held on present online databases). Turing covers 180 journals and reports from over 260 institutions. The database is also strong on conference literature. Turing should be available on Data-Star by late '88.

A more detailed analysis and advice on how best to search the various online bibliographical databases covering the literature of AI will appear in Hancox et al. [2].

REFERENCES

1. Mills, W.J., AI: Just how scattered is the literature? An online investigation. SIGART Newsletter 99 (1987) pp. 24-26.

2. Hancox, P.J., Mills, W.J. and Reid, B.J. - Keyguide to information sources on Artificial Intelligence and Expert Systems. London: Mansell ISBN 0-7201-2007-1 (Forthcoming 1989).

The Infinite State Acceptor And Its Application to AI

Edward A. Ipser, Jr.
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
ipser@vaxa.isi.edu

Computer Science Department
University of Southern California
Los Angeles, CA 90089-0782

Abstract

The infinite state acceptor is presented and is shown to be more powerful than the Turing machine for purposes of accepting strings. The role of abstraction in the design of such machines is discussed. It will be argued that such an acceptor is not merely an eccentric extrapolation of current models but is useful for the description of the behavior of devices whose complexity is non-trivial.

Introduction

Loosely speaking, a finite state acceptor is a machine with a finite number of states (and state transitions) and no memory. A Turing machine is a machine with a finite number of states but with an infinite amount of memory.

However, strictly speaking, computers, being finite, could most accurately be modeled as finite state acceptor rather than as Turing machines.

> To be sure, we will always be confined, in real life, to machines which are finite. But I assert that it is not always the *finiteness* of the machines that limits their uses; more usually it is either (1) the *practical limitations of running-time* or (2) the conceptual complexity of their structures or "programs".[1]

To take this kind of argument a little futher, suppose that we drop our insistence upon finite descriptions of behavior; it is not clear that we are limited to machines of finite complexity.

The current trend in software design is to generate systems with properties which can be formally tested and proven. The objective of this paper is to justify an alternative approach to the design of machines, especially intelligent machines. Specifically, i propose that insistence upon algorithmic devices is not only constaining but self defeating. Instead, the trend of this approach is quite the opposite; the systems that are generated will be unpredictable in the sense that a finite specification cannot predict the behavior of the system.

The Infinite State Acceptor

The infinite state acceptor is a simple extension of the finite state acceptor. As its name implies, it is simply a lifting of the restriction that a machine have a finite number of states.[1]

The infinite state acceptor is defined as follows:

Definition: An *infinite state acceptor* (abbreviated *ISA*) is a 5-tuple $Z = (K, \Sigma, \delta, p_0, F)$ where

- K is a (possibly infinite) set (of *state symbols*);
- Σ is a finite set (of *input symbols*);
- δ is a function from $K x \Sigma$ into K (the *next state symbol* function);
- p_0 in K (the *start state symbol*); and
- F is a subset of K (the set of *accepting state symbols*).

The configuration or snap shot of the execution of a infinite state acceptor consists of the current state and the unread input. At each step, one input symbol is read and discarded.

Definition: The *configuration* of an infinite state automata is a 2-tuple $C_Z = (p, w)$ where:

- p (the current state) is in K; and
- w (the remaining input) is in Σ^*.

Definition: The *next configuration function* ($\gamma(C)$ where $C = (p,w)$) of an infinite state acceptor is defined as follows:

- if $w = \varepsilon$ then $\gamma(C) = C$; else
- $\gamma(C) = \gamma((\delta(p,a),x))$ where a in Σ and $w = ax$.

The set of words accepted by an infinite state acceptor is defined as follows:

Definition: For each ISA $Z = (K,\Sigma,\delta,p_0,F)$ let $T(Z) = \{w \mid w \text{ in } \Sigma^* \text{ and } \gamma(p_0,w) = (q,\varepsilon) \text{ and } q \text{ in } F\}$. A word w is said to be *accepted* by Z if w is in T(Z) and T(Z) is said to be the set of words *accepted* by Z.

[1] The restriction that a machine have a finite vocabulary is also lifted but this is not necessary and next state function, will, of course, no longer be a finite mapping but this is nothing new; any mathematical function from one infinite set to another may have an infinite mapping.