Op weg naar een intelligente Grid

ENS

- Enterprise Nervous System -

- versie 2.03 -

Inhoudsopgave

Over de bijlagen

De eerste twee bijlagen zijn artikelen over Cyc waarbij **CycSecure** in de tweede bijlage centraal staat. De derde bijlage is een beschrijving over de structuur van Cyc. De laatste bijlage gaat over de doelstelling educatie voor alle wereldbewoners geschreven door Alfred Bork.

1 Op weg naar een intelligente Grid

Hedendaagse netwerken missen het vermogen om intelligent zaken af te kunnen handelen. Dit kunnen reparaties, ontwikkelen van processen en deze te implementeren of communicatie met de buitenwereld zijn. Wat reparaties betreft; netwerken kunnen fouten opvangen maar niet geheel herstellen of een situatie veranderen. Dit is moeilijk te realiseren omdat de standaard hardware die nu wordt ingezet zich daarvoor niet leent en het een zekere intelligentie mist. Wanneer we kijken naar reparaties zijn er verschillende herstel niveaus. Sommige reparaties hoeven niet met intelligentie te worden afgehandeld (zoals bij mensen, een kleine beschadiging aan de huid repareert het lichaam zelf). Genetische algoritmen zouden hiervoor voldoende zijn, een bepaald redenerend vermogen is niet nodig. Een intelligent netwerk gaat nog een stap verder. Intelligente netwerken kunnen zich aanpassen en daarbij zichzelf herontwerpen. Normaal levert de mens deze intelligentie maar als we nu eens een kunstmatige gezond verstand intelligentie inzetten zoals Cyc? Ik denk namelijk dat Cyc (intelligentie) en de Hypercomputer (adapteerbare hardware) waardige componenten zijn voor het upgraden van het Grid naar een intelligent netwerk [1, 2] die ondermeer voor het bedrijfsleven, onderwijs en wetenschap ten dienste kan worden gesteld en het doel *onderwijs voor iedereen op de wereld* behaald kan worden.

1.1 De technologie, als eerste de Hypercomputer

"Starbridge has created a powerful reconfigurable FPGA-based accelerator board that will run compute-intensive code orders of magnitude faster than is possible on traditional distributed Linux clusters. This significant speed advantage is due to the highly parallel nature of FPGA hardware, which is exploited by specifically tailored parallel code written using the Starbridge FPGA development environment, Viva®. Instead of performing one calculation per clock cycle per cluster CPU, the Hypercomputer can perform tens of thousands of calculations per clock cycle on a single accelerator board."

"The acquisition cost of the HC-62 Hypercomputer is comparable to a PC cluster of approximately 150 CPUs. However, it can deliver the same seismic imaging performance as more than five thousand Intel CPUs."

"The Hypercomputer fits into a single 4U cabinet, greatly reducing space requirements. It drastically reduces power consumption–only five hundred watts of power, in comparison with the more than five hundred thousand watts of power consumed by four thousand CPUs. Accordingly, air conditioning requirements are proportionally reduced. System administration effort is reduced, and system reliability is increased proportionally to the number of systems employed."

-- Seismic Processing with Tricon's Tsunami Suite Accelerated by Starbridge Hypercomputing (zie de attachment TRICON-ESS-Starbridge-Whitepaper.pdf)

Hypercomputers [3] kun je ook Field Programmable Gate Arrays (FPGA) computers noemen, aangezien dat de hardware basis is. Om deze computers snel te kunnen programmeren gebruikt men de 'compiler' Viva. Viva compiled eigenlijk niets, maar

kneed de FPGA in de juiste configuratie, met specifieke algoritmes, waarna er heel snel gerekend kan worden. Viva let ook de gesteldheid van de FPGA's en de data. Wanneer een component beschadigt raakt en niet meer goed functioneert wordt de data gered en op een andere FPGA verder verwerkt. Vandaar de term fault-recoverability. Fault tolerence is passé.

Een zwaar beschadigde Hypercomputer kan blijven functioneren, mits er stroom is. Dit is ideaal wanneer we wan een netwerk eisen die onder alle omstandigheden zijn taken kan uitvoeren en voltooien.

Wat de hypercomputer uniek maakt is dat de grens tussen software en hardware vervaagt. De FGPA kunnen direct on the fly hergeprogrammeerd worden voor de meest uiteenlopende taken. Als voorbeeld zie de attachment FPGA-protecting-networks.pdf (op deze manier van filteren van verkeer via dynamisch configureerbare hardware gaat vele malen sneller dan via software, vandaar dat iedere handeling, beter gezegd berekening, direct door de hardware dient te worden verwerkt i.p.v. door de software. Op de huidige manier van software toepassen leidt tot een grote overhead qua tijd en resource verbruik).

Naast de enorme rekenkracht is het spotgoedkoop vergeleken met geclusterde computers. De snelste Hypercomputer, de HC-124 van het formaat van een dubbele 4U kabinet, heeft evenveel rekenkracht als 10.000 geclusterde cpu's en verbruikt minimaal 300 watt en maximaal 1100 watt. De HC-124 kost $700.000,-. Dus wanneer de dollar blijft zakken t.o.v. de Euro wordt het alleen maar goedkoper. Het clusteren van 10.000 cpu's kost veel meer tijd en geld en dan vermeld ik nog niet eens het energieverbruik. Geclusterde Hypercomputers zijn helemaal prachtig. Want de rekenkracht neemt lineair toe met elke nieuwe Hypercomputer die aan de cluster wordt toegevoegd. Hypercomputers hebben in principe geen koeling nodig en kunnen in de meest extreme omstandigheden (het heelal kent extreme temperaturen schommelingen) worden ingezet.

Hypercomputers, of kleinere versie in de vorm van bijvoorbeeld routers (zie nogmaals attachment FPGA-protecting-networks.pdf), zouden handig zijn als backbone voor een intelligente Grid. De hardware kan on the fly worden aangepast, wat het netwerk adapteerbaar maakt voor de omgeving. Stel dat iemand een enorme schop geeft tegen het kastje, waardoor verschillende gedeeltes van de FPGA chips worden beschadigt, Viva zorgt ervoor de beschadigde delen niet meer worden gebruikt en routeert alle belangrijke functies naar de gezonde (gedeeltes van de) FPGA's. Nog een stapje verder en we kunnen er voor zorgen dat het netwerk zelfbeherend wordt, mogelijk met Cyc.

1.2 De technologie Cyc

Cyc [4] is een 'commen sense artificial intelligence'. Je zou het zelf redenerende software kunnen noemen. Cyc redeneert of denkt met behulp van miljoenen basisregels, wat ook gezond verstand wordt genoemd. Wanneer iets niet voor Cyc bekend is begint het er vragen over te stellen. Op deze manier kan Cyc ook aftasten hoeveel een bepaald persoon iets weet, en zich aan zijn/haar kennisniveau aanpassen. Cyc kan ook andere Cycs over bepaalde onderwerpen vragen stellen wanneer het zelf en de menselijke agent er niet mee bekend zijn.

Er zijn verschillende ‘versies’ of beroepen van Cyc. Zo is er CycSecure die instaat is om een geheel netwerk te beheren in bijvoorbeeld beveiliging en het bijhouden van updates etc. zonder menselijke hulp. Cyc kan gebruikers analyseren, en ook aanvallers (crackers) en al wat in aanraking komt met zijn netwerk bestuderen. Voor crackers kan Cyc een geheel virtueel netwerk creëren. De cracker denkt nog steeds dat op het echte netwerk inbreekt. Zonder dat de cracker het doorheeft houdt Cyc iedere handeling bij en deze wordt gelogd en verder geanalyseerd. Cyc kan de menselijke beheerder van advies voorzien en bij toestemming van de admin deze zelf implementeren. In principe is de mens in dit geval vervangbaar.

Cyc kan in meer gebieden worden ingezet, naast het beheren van netwerken.

Er zijn namelijk andere (soorten) Cycs. Verschillende Cycs verzamelen kennis waarbij ze het internet afstruinen opzoek naar nieuwe informatie voor de Cyc Knowledge basis gevestigd in Cycorp Inc. Deze kennis basis wordt geleidelijk gedeeld in de open bron versie van Cyc te downloaden op http://www.opencyc.com.

Doch, ook al weet Cyc veel, het is nog maar op het niveau van een 6 jarige qua menselijk inzicht. Cyc voldoet al wel om als een soort tutor te dienen. Een interactieve redenerende encyclopedie die zich aanpast aan het kennis en redeneringsvermogen van een mens die zo op zijn eigen tempo kan leren. De ontwikkelaars van Cyc denken dat de komende tien a vijftien jaar cruciaal zijn voor de ontwikkeling van Cyc naar volwassenheid. Meer hierover in de lezing over Cyc [5] (aanrader).

Cyc is denk ik de perfect interface voor het Grid netwerk [6], mogelijk in combinatie met hologrammen [7] voor gebruik met betrekkin tot communicatie. En zoals de opmerking was op 11 december 2003 (tijdens de oprichting van het Grid Forum Nederland) uit de zaal (van een wiskundige); het enige gebied waarop Nederlandse Grid groep nog echt kan scoren is een goede interface! Hologrammen kunnen een onslijtbare interface vormen.

Aangezien CycSecure netwerken kan onderhouden moet het ook mogelijk zijn om Cyc direct met de hardware, FPGA’s, te laten beheren. Het is een onderzoek waard om uit te zoeken hoe Cyc en de Hypercomputers kan worden geïntegreerd in het Grid netwerk.

1.3 Het probleem met standaarden

Wij mensen gebruiken grammatica als een standaard, maar gelukkig hebben we gezond verstand (al zou je dat soms niet zeggen) en kunnen we accenten, versprekingen en nieuwe woorden of grammaticale veranderingen opvangen. Netwerken kunnen dit niet aangezien zij een bepaalde vorm van intelligente missen. Als we echt het standaardprobleem willen oplossen dienen we een netwerk te gebruiken die kan redeneren. Dit intelligente netwerk kan met iedere nieuwe verandering of update overweg en wanneer nodig het oude accent of protocol gebruiken voor programma’s of personen die dit wensen.

Cyc biedt ons de kans om standaarden op een gemakkelijke manier te manipuleren naar onze wensen i.p.v. de nu huidige krampachtige houding t.o.v. standaarden; de één wil alles vast leggen, terwijl de ander het liefst alles zo soepel en open mogelijk

wil houden. Een netwerk dat begrijpt is handiger dan onze huidige niet intelligente netwerk zoals het internet en Grid netwerken.

1.4 Doel: een intelligent netwerk

Het doel is om een intelligente Grid netwerk te bouwen die volledig aan te passen is door ons en zichzelf kan aanpassen aan de omgeving. Het huidige Grid netwerk is internet op steroïden. Om een intelligent netwerk te kunnen verwezenlijken dienen de hardware en de software veel flexibeler te zijn dan ons huidige starre chips en programmatuur. Cyc en FPGA's, of te wel de Hypercomputer, bieden ons de kans om zo'n netwerk te verwezenlijken. Ik noem dit soort netwerk een Enterprise Nervous System, afgekort ENS. Dit netwerk kan iedereen ten dienste staan binnen ons zonnestels.

1.5 Context: Global Education for All

Everyone on earth, everywhere and at all ages, should be able to learn. We need education for all children in the world for the survival of humanity. This is essential for human happiness, for eliminating poverty, and for solving the world's major problems, such as population, water, and violence. A recent book by J. F. Rischard, High Noon, lists education for all as one of the twenty global problems that must be solved in the next twenty years.

-- Alfred Bork

Naast dat het ENS de wetenschap en commercie kan dienen is het ook mogelijk om een wereldprobleem op te lossen. 1 van de 20 wereldproblemen volgens vice-president van de Wereld Bank Jean-François Rischard [8] is educatie. Een gebrek aan goede educatie levert ondermeer een grote negatieve bijdrage in de vorm van armoede. Een intelligent robuust netwerk kan mogelijk een bepaalde bijdrage leveren qua oplossing met betrekking tot dit wereldprobleem. In het verlengde kunnen we andere problemen verzachten, zoals armoede.

Een intelligent netwerk heeft een intelligente interface die talen begrijpt, nooit moe wordt en altijd geduldig is. Dit is niet alleen wenselijk voor vele leerlingen, maar ook voor ouderen, voor iedere individu op deze planeet. Een leerstof die naadloos past bij een student behoort tot de mogelijkheid.

Zou het interessant zijn om Cyc en de Hypercomputer technologie eens onder de loep te nemen en onderzoeken hoe we dit in een ENS raamwerk kunnen toepassen?

Het beste, Joram Zutt
http://www.bazaarmodel.net

Noten

[1] Zie bijlage 2 voor een voorbeeld van intelligente communicatie met de buitenwereld en het zelflerend vermogen van het netwerk, in dit geval 'hoe Cyc leert'.

[2] In zijn PhD scriptie van Dr. Hugo de Garis etaleert hij in hoofdstuk 11 paragraaf 11.6.1 een idee hoe FGPAs kunnen worden gebruikt voor evolueerbare hardware. Dr. Hugo de Garis gebruikt Xilinx FGPAs in zijn evolueerbare neurtale netwerk structuur. Xilinx 2 FGPAs zijn chips die in de Hypercomputer worden gebruikt. Zie de bijlage **Bijlage-Paragraaf-11.6-en-11.6.1.pdf**

"Since ... FPGAs can be re-programmed an unlimited number of times, they can be used in innovative designs where hardware is changed dynamically."

Scriptie: http://www.cs.usu.edu/~degaris/papers/thesis/thesis.html
http://www.cs.usu.edu/~degaris/

[3] Starbridge: The Hypercomputing Company http://www.starbridgesystems.com/

[4] Cycorp Inc. http://www.cyc.com/

[5] **Computers versus Common Sense An Engineering Approach to AI**

Who:
Doug Lenat (Cycorp, Inc.)

Where:
Stanford University

When:
6/4/2003 4:15 PM (Pacific)

Length:
75 minutes

Series:
Seminar on Computer Systems 2002-2003

http://murl.microsoft.com/LectureDetails.asp?1032

[6] Kowledge Grid staat voor versie OpenCyc 0.9 gepland: http://www.opencyc.org/releases/

[7] **IO2 Technology** http://www.io2technology.com/ , **Zebra Imaging** http://www.zebraimaging.com/ en **FogScreen Inc**. http://www.fogscreen.com/ (Download or view New MPEG Video featuring the first two FogScreen prototypes in operation and the president of Finland walking through it in the InnoSuomi prize seremony (MPEG format. Viewing time 1 min. 30 sec, total 43 MB)).

[8] http://www.rischard.net/

Bijlage 1 AI's $60m question

Computer boffins pop AI's $60m question

June 09 2002 at 12:37PM
Quickwire
http://www.opencyc.com

Austin, Texas - Day after day since 1984, teams of programmers, linguists, theologians, mathematicians and philosophers have plugged away at a $60-million (about R600-million) project they hope will transform human existence: teaching a computer common sense.

They have been feeding a database named Cyc over a million truths and generalities about daily life so it can automatically make assumptions humans make: Creatures that die stay dead. Dogs have spines. Scaling a cliff requires intense physical effort.

Though some critics question the potential of this painstaking effort, the inventors believe Cyc will form the brains of computers with supercharged reasoning abilities - which could help us work more efficiently, make us understand each other better and even help us predict the previously unforeseeable.

Cyc (pronounced "psych") has already helped Lycos generate more relevant results on its Internet search engine. The military, which has invested $25-million in Cyc, is testing it as an intelligence tool in the war against terrorism. Companies use Cyc to unify disparate databases and are examining a new application that warns when computer networks have vulnerabilities hackers can exploit.

'We stand on the threshold of success'

This spring, the developers' company, Cycorp Inc, sent their 18-year-old creation off for some higher education, creating a Web link to let the public download Cyc's knowledge base and teach it things too.

Cycorp's founder and president, Doug Lenat, believes that if enough people log in to share more of the world's collective wisdom, Cyc quickly will become vastly more useful.

For now, Cyc is just a few hundred megabytes that can be stored on a single CD. Someday, Lenat envisions it becoming standard equipment in computers or being placed on a network server to fuel dozens of applications. It could annotate emails to put them in better context for their recipients, serve as an instant language translator, even offer humans advice from varying points of view.

"This is the most exciting time we've ever seen with the project. We stand on the threshold of success," Lenat, 51, said recently in Cycorp's offices in a quiet Austin complex. "What people are able to do on a day-by-day basis could be dramatically increased if we are successful."

Such hopes are not new in artificial intelligence, which has to date produced far more disappointment than marvel, with products of only limited use beyond specific tasks.

'Generalise as much as possible'

In 1983, when Lenat was a professor at Stanford University and a researcher for Atari, he decided artificial intelligence would go nowhere unless someone took the time to create a catalog of common sense that would let a computer recognize absurdities as well as humans can.

With colleagues at Microelectronics Computer Corp, a technology research consortium, Lenat began creating Cyc in 1984.

By typing messages in CycL a programming language created especially for Cyc, Lenat's team first taught it that there are things in the world, and that some are individual (such as the Parthenon) and others are collections (historic sites).

The programmers eventually took chunks of text and thought about every assumption the author knew readers would make. Upon reading something about how the Duke of Wellington was moved by Napoleon's death, the programmers decided to tell Cyc it could assume Wellington outlived Napoleon, knew him when he was alive, heard about his death - and so on.

The goal was not just to fill Cyc with straightforward facts but to "generalise as much as possible until further generalisation would be false," Lenat said.

The result is that if you ask Cyc whether Lassie has a nose, it would reason that Lassie is a collie, collies are dogs, dogs are macroscopic vertebrates and macroscopic vertebrates have noses, so yes.

The researchers also told Cyc to ask questions if it decides it needs more clarity about a concept.

In 1986 Cyc asked whether it was human. That same year it asked whether any other computers were engaged in such a project.

Lenat's team taught Cyc to make sure everything it was told conformed with everything it already knew - a protection that should keep Cyc from being filled with erroneous information during its public education, which for now is possible only on computers with the Linux operating system.

Cycorp was spun off in 1994 into a privately held company that Lenat says has been profitable from inception, funded by the government, private investors and side projects such as the Lycos search-engine deal, which ended last year.

Cyc's job at Lycos was to make sense of ambiguous search results. If a user entered "vets", Cyc would ask whether he meant veterinarians or veterans and then have appropriate follow-up questions.

Amusingly, the Lycos stint provided Cyc with an adolescence, because it learned about sex-related terms users typed into the search engine. Cyc's programmers taught it that certain things in the world are salacious and shouldn't be mentioned in everyday applications.

Some artificial intelligence experts question whether Cyc can be as revolutionary as Lenat predicts. They claim it is far more efficient to make computers search for and identify patterns than to have them follow predetermined sets of rules.

"There's a lot more to common sense than can be captured in a set of rules," said . Pandurang Nayak, chief architect for Stratify Inc, which uses pattern-matching technology to manage data in various kinds of files. - Sapa-AP

Bijlage 2 CycSecure in use

Cyc in use

Story by Gary H. Anthes

APRIL 08, 2002 (COMPUTERWORLD) - A top-secret military command, which asked not to be identified, is pickier than most organizations about information security. And it's one of the first users of CycSecure, the first commercial application of Cycorp Inc.'s Cyc Knowledge Base. The command uses CycSecure to ensure that all relevant patches have been applied to fix known vulnerabilities in its networked computers.

CycSecure taps into a variety of sources, including the U.S. Department of Defense's Information Assurance Vulnerability Assessment notification program, to keep up to date on all known ways a system can be attacked. It also knows about the military command's computers and networks and combines that knowledge with the vulnerability information to simulate network attacks. When it spots a potential vulnerability in a computer, it can go out to that box to determine whether it is in fact vulnerable and then recommend the appropriate patch.

"It can plan an attack from the outside or from a disgruntled employee on the inside," says an information systems security analyst at the command. "It's pretty amazing."

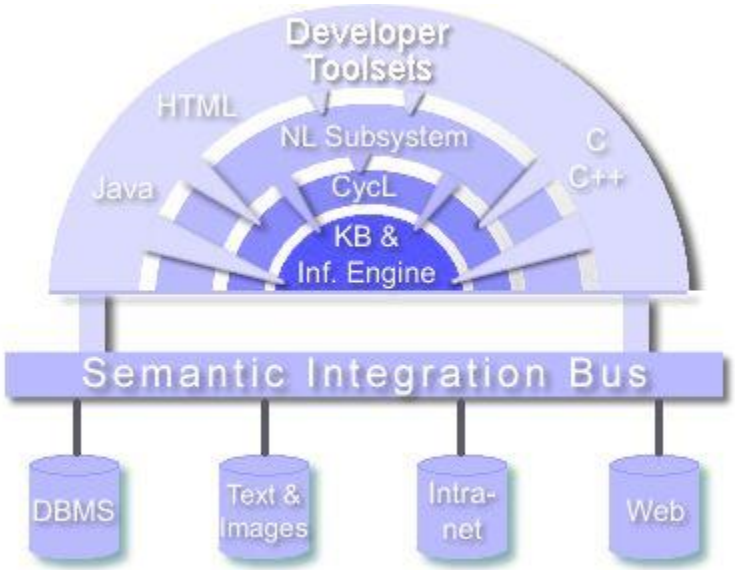
She says Austin, Texas-based Cycorp customized CycSecure for the command, but because such customization involves just adding rules and knowledge to the database, it doesn't require software changes. Cycorp maintains the application for the command now, but eventually the user will be able to take over maintenance itself, she says.

"It's rule-driven; it's dynamic; it kind of grows," she explains. "It keeps up with the attacks that are out there."

Cycorp CEO Doug Lenat offers this explanation of CycSecure: "Cycorp Cyc knows what are normal, legitimate actions -- such as a user renaming one of their own files or changing their password -- and what are actions taken by hackers -- such as packet-sniffing and spoofing. An attack plan generally includes a large number of 'normal' steps and one or more 'hacker' steps. Cyc does not have a model of the hacker mentality, such as goals, ego and so on, but it does have the notion that hackers generally want to be undetected, since that motivation accounts for many steps in many plans which would otherwise be missed."

Bron:
http://www.computerworld.com/industrytopics/defense/story/0,10801,69848,00.html

Bijlage 3 The Cyc Knowledge Server



The Cyc Knowledge Server

The Cyc Knowledge Server is a very large, multi-contextual knowledge base and inference engine developed by Cycorp. Cycorp's goal is to break the "software brittleness bottleneck" once and for all by constructing a foundation of basic "common sense" knowledge--a semantic substratum of terms, rules, and relations--that will enable a variety of knowledge-intensive products and services. Cyc is intended to provide a "deep" layer of understanding that can be used by other programs to make them more flexible. The Cyc technology includes a number of essential components:

- The Cyc Knowledge Base
- The Cyc Inference Engine
- The CycL Representation Language
- The Natural Language Processing Subsystem
- Cyc Semantic Integration Bus
- Cyc Developer Toolsets

The Cyc Knowledge Base

The Cyc knowledge base (KB) is a formalized representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life. The medium of representation is the formal language CycL, described below. The KB consists of terms--which constitute the vocabulary of CycL--and assertions which relate those terms. These assertions include both simple ground assertions and rules. Cyc is not a frame-based system: the Cyc team thinks of the KB instead as a sea of assertions, with each assertion being no more "about" one of the terms involved than another.

The Cyc KB is divided into many (currently hundreds of) "microtheories", each of which is essentially a bundle of assertions that share a common set of assumptions; some

microtheories are focused on a particular domain of knowledge, a particular level of detail, a particular interval in time, etc. The microtheory mechanism allows Cyc to independently maintain assertions which are *prima facie* contradictory, and enhances the performance of the Cyc system by focusing the inferencing process.

At the present time, the Cyc KB contains tens of thousands of terms and several dozen hand-entered assertions about/involving each term. New assertions are continually added to the KB by human knowledge enterers. The aforementioned numbers do not include (i) non-atomic terms such as (`#$LiquidFormOf #$Nitrogen`), nor (ii) the vast number of assertions added to the KB by Cyc itself as a product of the inferencing process.

Inferencing in Cyc

The Cyc inference engine performs general logical deduction (including modus ponens, modus tolens, and universal and existential quantification), with AI's well-known named inference mechanisms (inheritance, automatic classification, etc.) as special cases. Cyc performs best-first search over proof-space using a set of proprietary heuristics, and uses microtheories to optimize inferencing by restricting search domains.

Because the Cyc KB contains hundreds of thousands of assertions (aka "rules"), many approaches commonly taken by other inference engines (such as frame-based expert system shells, RETE match, Prolog, etc.) just don't scale up to KBs of this size. As a result, the Cyc team has been forced to develop other techniques.

Cyc also includes several special-purpose inferencing modules for handling a few specific classes of inference. One such module handles reasoning concerning collection membership/disjointness. Others handle equality reasoning, temporal reasoning, and mathematical reasoning.

CycL: The Cyc Representation Language

CycL, the Cyc representation language, is a large and extraordinarily flexible knowledge representation language. It is essentially an augmentation of first-order predicate calculus (FOPC), with extensions to handle equality, default reasoning, skolemization, and some second-order features. (For example, quantification over predicates is allowed in some circumstances, and complete assertions can appear as intensional components of other assertions.) CycL uses a form of circumscription, includes the unique names assumption, and can make use of the closed world assumption where appropriate.

For more information, see the detailed description of CycL language features.

Natural-Language Processing

Natural-language (NL) processing is among the most studied -- and most intractable -- outstanding challenges of software engineering. Many teams have attempted to produce NL systems capable of reading and making sense of plain english text, but none have succeeded to any significant degree outside of narrow, pre-conceived domains. As shown in the examples below, CYC-like common sense is a prerequisite for human-level competence at this task.

Consider the following pair of sentences:

- Fred saw the plane flying over Zurich.
- Fred saw the mountains flying over Zurich.

Although the sentences are very similar, humans have little difficulty in recognizing that in the first sentence, "flying" probably refers to the plane, while in the second sentence, "flying" almost certainly refers to Fred. Traditional NL systems will have difficulty resolving this syntactic ambiguity, but because CYC knows that planes fly and mountains do not, it will be able to parse these sentences just as easily as a human. It's difficult to see how this could be done without relying on a large database of common sense.

Here are a couple more examples; these involve pronoun disambiguation:

- The police arrested the demonstrators because *they* feared violence.
- The police arrested the demonstrators because *they* advocated violence.
- Mary saw the dog in the store window and wanted *it*.
- Mary saw the dog in the store window and pressed her nose up against *it*.

The CYC-NL system has three components: the lexicon, the syntactic parser, and the semantic interpreter. These are described in greater detail on the CYC-NL page. At the moment, we are focusing our efforts on broadening the coverage of all three components of CYC-NL. We are currently able to correctly parse many different sentence types, including ambiguous and syntactically complex inputs. CYC is capable of handling negation, modals, and nested quantifiers. We are developing interfaces which will allow people to make assertions and query CYC using english instead of CycL. We also are working on a generation component, which will produce english strings from CycL formulas.

CYC's NL capabilities form the foundation for applications in knowledge-enhanced searching of captioned information, and for user-friendly interfaces to other applications, including the database integration application.

Future directions for CYC-NL will include:

- exploring the role CYC could play in machine translation
- using CYC-NL to post-process output of speech recognition systems
- harnessing CYC-NL to enhance user interfaces.

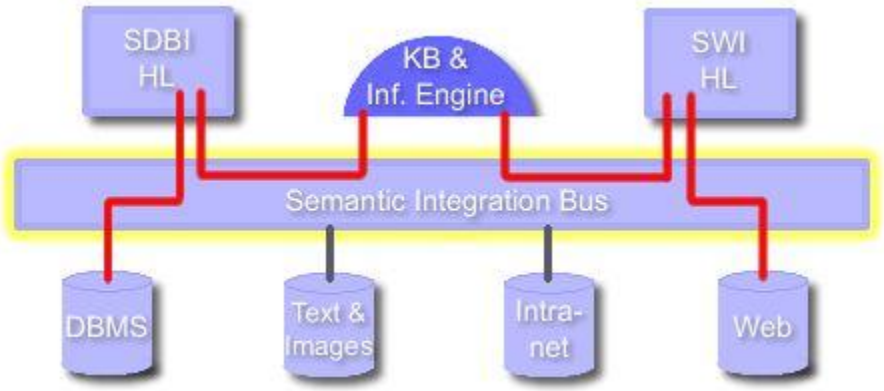
Other potential applications are myriad.

For more information, see the more detailed description of Cyc NL subsystem.

Semantic Integration Bus

Computer-based information is stored in many forms, including data that is structured (databases), semi-structured (spreadsheets, web pages), and unstructured (text files and text fields). Cyc can turn some of this information into usable knowledge, and the remainder can be annotated for easier access by humans.

Cyc treats each database record as if it were an implicit assertion in the knowledge base. These implicit assertions are then available during inference. Similarly, text fields can be read using the natural language processor to see if they contain any useful implicit assertions. Sometimes the assertions describe what the text is "about". Cyc can use this information to locate and report information resources which the user may employ to answer a particular query.



In the above diagram, information stored in a database or on the web is made available to the inference engine as virtual assertions. These sets of virtual assertions are managed by heuristic level (HL) modules. For example, the inference engine "broadcasts" a query on the bus. An HL module recognizes that the request asks for an assertion which maps into its virtual knowledge space. The HL module intercepts the request, communicates with the database, web site or other knowledge source, and returns bindings to the inference engine. Inference then continues, combining information from multiple sources.

Developer Toolsets

The Cyc system also includes a variety of interface tools that permit the user to browse, edit, and extend the Cyc KB, to pose queries to the inference engine, and to interact with the natural-language and database integration modules.

The most commonly-used tool, our HTML browser, allows the user to view the KB in a hypertexty way. HTML pages describing Cyc terms are generated on the fly by the Cyc system. Each page describes a Cyc term by showing all the assertions in which it is involved, organized according to a standard schema. Every occurrence of a Cyc term is an HTML link to a (dynamically-generated) HTML page describing that term, so that it is easy to surf around the KB following a network of relationships. The HTML browser also includes facilities for searching and editing the KB and for posing queries to the inference engine.

Other HTML interface tools include:

- A "KE file" tool, which allows users to batch-process text files describing new material to be added to the KB.
- A hierarchy browser, which displays any desired subtree of the Cyc subset tree in outline format.
- A lexicon editor, which provides a user-friendly way to edit and extend the Cyc lexicon.
- An English-to-CycL parser, which lets users experiment with Cyc's natural language facilities by parsing arbitrary English strings.
- A database tool interface, which provides an interface to Cyc's semantic integration module.
- A WordNet browser, which allows users to view WordNet in relation to the Cyc ontology.
- An English generator, which restates rules in Cyc in English.

Bijlage 4 Attaining Education for All

ATTAINING EDUCATION FOR ALL

Alfred Bork
School of Information and Computer Science
University of California
Irvine, CA USA

bork@uci.edu

January 23, 2003

> *Everyone agrees that the single most important key to development and to poverty alleviation is education.*
>
> James Wolfensohn, The World Bank

Introduction

Everyone on earth, everywhere and at all ages, should be able to learn. We need education for all children in the world for the survival of humanity. This is essential for human happiness, for eliminating poverty, and for solving the world's major problems, such as population, water, and violence. A recent book by J. F. Rischard, High Noon, lists education for all as one of the twenty global problems that must be solved in the next twenty years.

A series of international conferences have addressed the question of Education for All. They set a series of initial goals to attain this desired state for learning in the world. But these have seldom been met. The latest conference, in Dakar, set goals to be reached in 2015. The three major goals were

- Primary education for everyone
- Equal educational treatment of boys and girls.
- Cut illiteracy by half in the world

These goals are only the first step in providing education for everyone on earth. The idea is to approach the problem gradually. Various organizations are involved.

> The EFA partnership, which is a broad coalition of national governments, organizations such as UNESCO and the World Bank, civil society groups, and associations, is committed to reaching and sustaining the EFA goals through broad-based partnerships within countries, and supported by co-operation with regional and international institutions.

The World Bank

A UNESCO study released on November 13, 2002, examined the progress of the world toward these goals. The results are discouraging. The evidence indicates that seventy countries will not reach these goals, in spite of massive expenditure by the World Bank and others. A report of this study is at
www.unesco.org/bpi/eng/unescopress/2002/02-93e.shtml

This paper asks why we are not reaching these goals. Then it considers a strategy not tried globally, but very promising, for attaining Education For All.

The Difficulties of Attaining Education For All

Why is it so difficult for the world to realize almost universally supported goals? There seem to be several key factors.

Numbers of people

We now have more than six billion people on earth. Population is projected to rise to nine billion by 2050. Most of this growth is in the poorest areas of the world, the places where education is the poorest. Most of our current strategies for learning come from a time when there were far fewer people on earth, and we have not coped well in education with increased numbers.

Unequal distribution of access to resources

The world is very uneven, not a level playing field. Some live at a high level, some at a very low level. Often schools do not exist, or are primitive. Some of these refer to cyberinfastructure.

Existing organizations are slow to change

This is a point stressed in High Noon, due to the top down structure of most organizations, such as schools and universities.

Poverty

Over 24,000 people, mostly young children, die of starvation each day. Many people live in very poor circumstances, lacking adequate shelter and drinking water. Kofi Annan reports that half the people on earth live on less that $2 a day. Although we think of these conditions as developing world circumstances, even developed countries are not immune.

Infrastructure

One third of the people on earth have never used a telephone. The vast majority has no connection to the Internet, and even if they had such a connection could not use their native languages.

Strategy

The major paradigm used so far to attain Education For All is to build schools and train teachers. But this may not be an adequate strategy for the present and the future, particularly in the developing world. This possibility is often ignored.

> I do not believe that the global education problem is solvable by conventional means: the building of classrooms in remote areas, and the preparation of a vast array of teachers. Very affluent countries may continue to pursue education this way, but even we in America have, almost without realizing it, created a vast system of nonconventional continuing education that today serves to educate more people than the conventional system, ranging from kindergarten to the university. We have certainly come to the time when we need to entertain some new and creative thoughts about the total enterprise of education, especially as it affects the less developed countries, which will become comparatively less and less developed without some new system of education.
>
> Theodore Hesburgh
> The Human Imperative, page 65
> Yale University Press, 1974

We should give serious attention to alternatives to schools and teachers that might help reach Education For All. This is the direction the remainder of this paper will pursue. There is, I will argue, a promising alternative to schools.

Key Questions for Alternative Learning Systems

In considering present and new systems for universal education we can raise these questions.

- Is the system sustainable for the foreseeable future? Often approaches depend on continuing financial support, or on special circumstances.
- Is the system scalable? Can it reach everyone on earth, at all ages? The Dakar goals refer mostly to young children, but they are not the full story.
- Is the system affordable, by the world, by the country, by the individual? This question is closely related to the first two. It suggests careful detailed evaluation of learning systems by economists, including current systems.
- Does almost everyone learn in this system? Success in learning is essential for encouraging further learning.
- Can learning be done on a flexible schedule? Students may have constraints on time available for learning, depending on their circumstances.
- Is learning motivating to students? We want a world in which people are eager to learn. We need to keep people at difficult learning tasks because they enjoy learning. We want people to learn even if their living situations are poor. We want learning to be lifelong.
- Is learning culturally sensitive? Do we address the values of each culture, including language? Young children should learn first in their native languages.

- Is learning potentially available everywhere? This if related to the issue of scalability.
- Does learning treat each person as a unique individual? Does it react frequently to individual student difficulties? Is learning personalized? Is it always helpful and encouraging?

A Suggested Alternative System

We have been proposing a system that may offer positive answers to all these questions. It is fully described in a book (Bork, Gunnarsdottir 2001) and in many papers at www.ics.uci.edu/~bork . Other information is available from the author.

The paradigm for this approach is a classical form or learning, with a skilled tutor and a very small group of students, no more typically than three. An early example is Socrates, as pictured by Plato in the early dialogs. Apprentice learning, even older, also furnishes clues.

Tutorial learning has been the preferred way of learning for those who could afford it. Thus very wealthy families often had and have private tutors for their children. But good human tutors are expensive, and limited in number. Home schooling can be a variant of this tutorial approach when there are knowledgeable parents that can devote large amounts of time to their children.

A new approach is possible now. The computer can now serve as the tutor. The following characteristics can be achieved.

- The material will be highly interactive, in the sense of Socratic tutorials.
- Questions from the computer will be frequent, typically at intervals no greater than twenty seconds. This figure is from research at the University of California, Irvine. The questions are developed by skilled teachers in the design groups.
- Answers will be free-form, in the student's native language. Very little pointing and clicking or multiple choice will be used. We need to use the full power of our languages, our most important intellectual tool, to make the decisions listed below.
- Voice will be used in two directions for most of this work. The computer will talk to the student, and the student will talk to the computer. Unrestricted voice is our best way to communicate, allowing a natural interaction. Current software is adequate in some languages, and can be extended to others. Training for individual voices will mostly not be needed, and where it is it will be combined with the learning material.
- The program analyzes the replies, and decides what learning material to present next, again as determined by the design groups and the evaluations. Stored student records may play a role in these decisions. The Vygotsky idea of the zone of proximal development will be useful.
- In each area the interaction continues until success is reached. Learning time may vary. To accomplish mastery for all, different learning material may be required for different students. Success is determined by the designers, reflecting benchmarks from many countries. In a very few situations human help may be suggested by the computer.
- A major purpose of interaction is to identify student learning problems as soon as possible after they occur, and offer effective assistance. The skilled

teachers in design groups pool their knowledge of likely problems and how to help with them.
- Mostly students work in groups of about three of four at the computers, allowing valuable learning with peer interaction. Occasionally students will be asked to work alone. Groups may be changed by the computer to allow for different rates of progress and to allow different students to work together. In addition to learning benefits, this approach will stimulate focused social interaction.
- Interaction with teachers, parents, and other adults will be possible.
- The experiment (mentioned next) will use present computers and CD or Internet access. For later widespread use less expensive computers perhaps with solar panels will be possible, and satellite distribution is likely.
- In developing and using the material careful information on costs will be collected. The important data is the cost for a student hour of high quality learning, considering all factors. We expect these to be less than current figures, with the sizable numbers of students we contemplate.

Production of such material demands a new strategy. We have been developing such a strategy, with supporting software, for over thirty years at the University of California, Irvine (Bork et al, 1992) (Bork, Gunnarsdottir 2001), with the help of people from the University of Geneva and California State University, San Marcos. The critical role of excellent teachers in the design has already been mentioned. Evaluation and improvement cycles are also critical.

Steps to Attain this Highly Adaptive System

While this approach is very promising, current limited experience is not enough. We suggest a multi step approach, focusing on the Dakar goals initially.

- Experiments

We need a better empirical basis for this work, at all levels of education. A major experiment with young children could determine if this approach can bring Education For All. We are proposing such an experiment, involving three areas: Reading and writing, mathematics, and scientific literacy. Some material would be initially developed in several languages. Extensive evaluation would include very poor students in regions with no schools. The cost, including evaluation, would be about $17 million. Details are available. Other experiments should cover other areas such as universities and training.

- Full development

If the experiments are successful, the world can proceed to full development of the new learning system, for all levels from birth to old age and for all subjects. This will be a major global human adventure. Careful thought needs to be given to organizing this development. Funding is also an important issue.

- Research

As large numbers of students use the new learning materials, the computers can save extensive data on student performance. This will allow much better empirical information than is possible now.

The New System in Practice

The following was written for a panel at ICCE 2002 in New Zealand. It illustrates what we believe is possible, if we start soon.

> Let me tell you a story. The time is 2025, in a town in India with about 100,000 people. We are in the center of town, in a room that contains many screens that look like television sets without keyboards. Many people all ages are intently interacting with the material on the screen. The interaction is entirely by voice, and the native language of the students is used, except when they are learning another language. Students are free to say anything they wish, and they do.
>
> Their slightly older brothers and sisters bring three five-year old students to the center, Maya, Asha, and Gopal. This is their first visit to the Center.
>
> They sit down at one of the screens, and begin a conversation with the 'learning aid'. The human voice is the mode of this interaction, in both directions. The learning device talks to the students, mostly asking questions, and the students talk to learning device, answering the questions. Our three young heroes sit down together, and often discuss what they will say.
>
> The learning aid first asks the children for their names. Then it suggests that Maya tell a story. It may need to prompt the student to begin, but probably this will not be necessary, as children love to tell stories. As the child tells a story print appears on the screen in the child's native language. Since these young children they may not know how to read the print. Pictures illustrate the story as is told. The learning device adds periods where appropriate. We are beginning even in this early stage both to get children to write, and to understand the connection between speech and the funny symbols that appear on the screen.
>
> The learning device asks the student to read back the story, soon determining just what they can read. The learning aid reads back the story slowly, using the voice of the student, emphasizing each word on the screen as it is read. The story is stored, and will be presented again to the student.
>
> Our students come back often, eventually learning to read and to compose stories and other documents. The devices learn much about the student, and use that information in future learning. Occasionally the learning device suggests that they work with other students, suggesting which students, so they become familiar will many children in the village.
>
> Other groups around the room are working on different things. Some learn about nonviolent ways to resolve disputes. A group of 10-year-olds is learning how to differentiate a function. An older group of children is working on quantum electrodynamics. Some are learning how to build and operate a small factory. Several 80 year olds are striving to understand Beethoven's ninth Symphony.

No one is in charge in the room. It belongs to the learners. Learners continue working on a topic, perhaps in several sessions, until they succeed in learning, as determined by the learning device. Since all succeed, no tests and grades are given. Any subject is possible at any time and at whatever pace the students want or need. The learning aid assists with choosing new subjects, based on what is known about the student.

The learners are not required to work in groups, but most do, in groups of about three or four. They come whenever they want to, and leave whenever they want to; they are free learners, working because they are in a highly interactive motivating environment, and because they have always enjoyed learning. Many of the children spend much of the day in the center. They may also work on projects outside the Center. If special help is needed, the learning aid suggests several people to assist, and asks those people if they are willing, when they are in the Center.

The learners in this village understand that learning is a lifelong process. They love to learn, perhaps because they are always successful in learning, and perhaps because they do it with their friends. Even in a young age they understand the power of learning, and this encourages lifelong learning.

They know too that billions of other people on the earth are engaged in similar learning activities, and they understand clearly that this helps them avoid violence, and to solve the major problems of the world, such as too many people, poverty, and not enough food or water.

Their own village has changed greatly since the learning aids were available. Everyone has a comfortable place to live, enough food, and enough water. People enjoy life.

References

Bork, Alfred, Bertrand Ibrahim, Alastair Milne, Rika Yoshii, The Irvine-Geneva Course Development (Bork et all 1992), in R. Aiken (Ed.), Education and Society, Information Processing 2, Elsevier, Holland, 1992.

Bork, Alfred (Bork 1999), Interview, EDUCOM Review, July/August 1999.

Bork, Alfred (Bork 1 2001), Tutorial Learning for the New Century, Journal of Science Education and Technology, March 2001, Volume 10, Number 1.

Bork, Alfred (Bork 2000), Four Fictional Views of the Future of Learning, The Internet and Higher Education, 3 (2000).

Bork, Alfred (Bork 2 2001), What is Needed Effective Learning on the Internet, Educational Technology and Society, 4(3) 2001.

Bork, Alfred, Sigrun Gunnarsdottir (Bork, Gunnarsdottir, 2001), Tutorial Learning – Rebuilding our Educational System, Kluwer Academic Publishers, New York, 2001.

Rischard, J. F., High Noon – Twenty Global Problems, Twenty Years to Solve Them, Basic Books, NY, 2002.

Papers are available at www.ics.uci.edu/~bork.