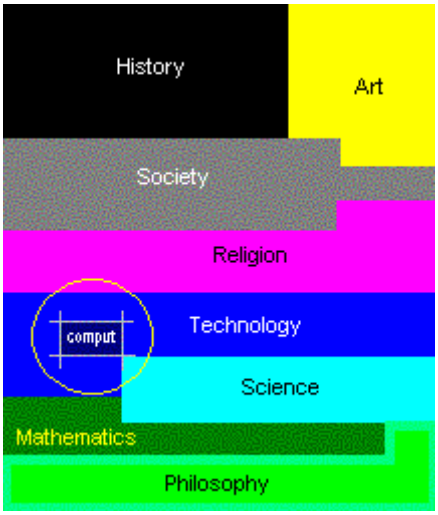
# *The Tree of Knowledge*

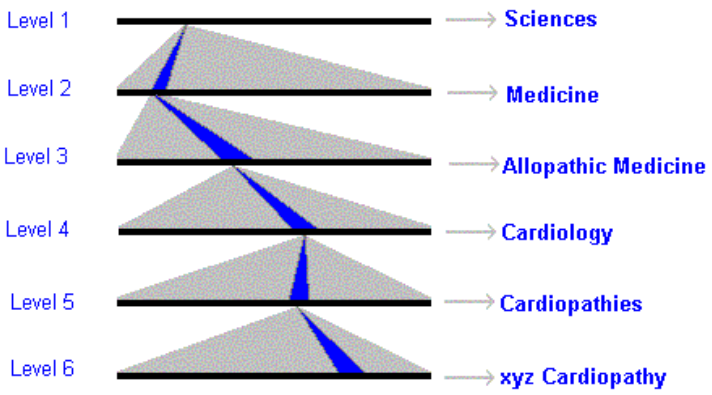
## Human Knowledge Maps

By Juan Chamero, CEO of Intag, as of May 2004

The figure below depicts a Knowledge Map at its first level, with its classical and ancient division in eight Major Disciplines. Each discipline curriculum could be in its turn thought hierarchically structured as a Logical Tree in such a way that making click with their mouse on it Internet users may go deep inside up to five or six, eventually seven levels, guided by a wizard.




Within Technology in a second level appears “Computing” which was entirely mapped to demonstrate the feasibility of the whole map (see it working at http://www.intag.org, an Open Source Website). In this case the selected curriculum was a combination of the American ACM 2002 Curricula and of the IFIP-Unesco Computing Curricula for the Rest of the World. The combined computing tree was opened in about 1,600 subjects split in four levels. We show below how a user is guided to look for a given xyz Cardiopathy in Medicine within Sciences.



For each subject the map presents a set of specific “keywords” that is those semantic particles that originate in that specific subject and for that specific level. The whole ordered collection of keywords for a given discipline, related by this criterion of specificity to its Logical tree, defines its Thesaurus. The Computer Thesaurus has more than 53,000 keywords.

A total Human Knowledge Map would have from 150 to 250 disciplines, depending of the subdivision criteria. Our estimation is 200 disciplines that split in 200,000 subjects with a Thesaurus of nearly 10 million keywords. As maps point to things, these knowledge maps should point to “cognitive” things, namely documents and or solutions and or “how to obtain some specific piece of knowledge code”. In order to satisfy users needs in terms of information these maps could be settled to point to a set of “reasonable good” cognitive objects for each level of search, namely for each path along the whole “Tree of Knowledge”. These cognitive sets closely related to subjects will resemble the fruits of the Tree of Knowledge.

These fruits are actually known as “authorities”, for instance documents that “teach” humans the “authorized best” about a given subject. If these documents deal with their respective subjects covering their whole specific spectrum, they could be considered “e-authoritative books” or simply “e-books” of the Human Knowledge Map. We may think of a virtual library architecture that has in the average non repeated (different approaches/points of view) from four to seven e-books per subject, making a volume of 1,000,000 e-books.

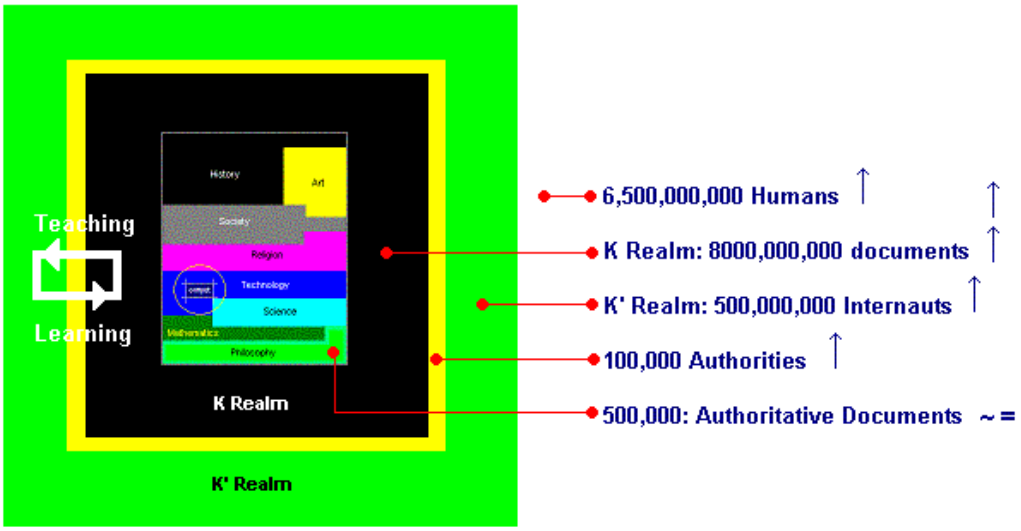
This Map of Knowledge/Tree of knowledge could be continuously updated by agents following the knowledge evolution. These maps are apt to offer at any moment the actual best of the human being in only one click. However users may feel unsatisfied with the fruits offered by grant. In those cases agents may search for them “similar” authorities in the “neighborhood” of the subjects’ tree up to any level of detail until exhaustion of available data. The neighborhood is probabilistically defined taking into account the “fingerprint” or semantic spectrum of fruits, their keywords that are equivalent to fruits’ seeds. Agents look for “similar” documents defined as those that have similar seed sets.

Users may query for subject, keyword or by pairs keyword, subject if they prefer so. They should take into account that a keyword may not define a document, and conversely a document usually deals with more than a single keyword.

## *Knowledge Maps Synthesis*

Our Darwin-FIRST technology emerges from a new paradigm of the Digitalized Knowledge, a new way to see the man-machine interaction in the Cyberspace. The Web could be seen as a giant interface between two Realms, the K Realm where the “established knowledge” is hosted interacting with the K’ Realm where the people as users interact. The established order “teaches” meanwhile the people “learn” as much and as efficiently as possible. Both realms look like alive governed by a Teaching-Learning cycle as a non separable unity.

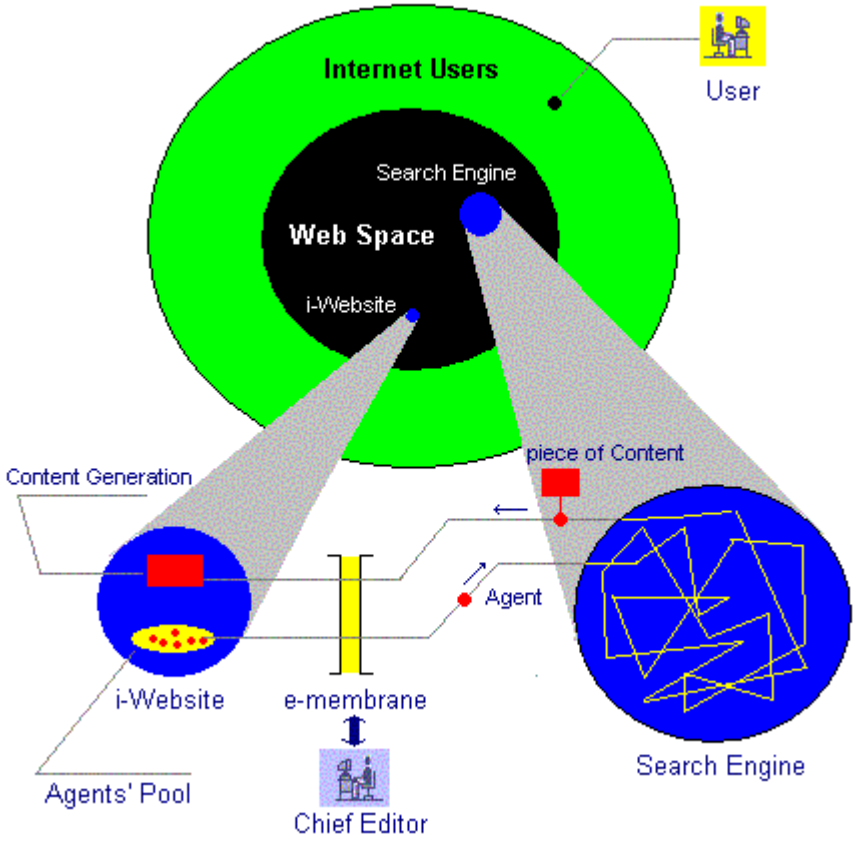
By K Realm we are actually talking of a content of nearly 8000 million documents (unfortunately by addressing problems only half of this figure is actually indexed by the conventional Search Engines). In the K’ Realm nearly 500,000 “internauts” are continuously surfing the Web, trying to satisfy their curiosity and information needs by “playing” versus the K Realm. Out from these 8000 million documents only a small collection of about 100,000 could be considered “authorities”. Within the Web space we may imagine a Human Knowledge Map indexing the whole K Realm, pointing directly to a selected core of about 500,000 Authoritative Documents, and through it to the whole content via neighborhood similarity.




This map could be cloned and tailored at human will. Maps evolve continuously in content and structure as driven/pressed by people's interactions. The core volume of Authoritative Documents remains pretty much constant along the time, some documents will die and some others will born, some disciplines turn obsolete or disappear, some others appear.

# *E-Membranes*

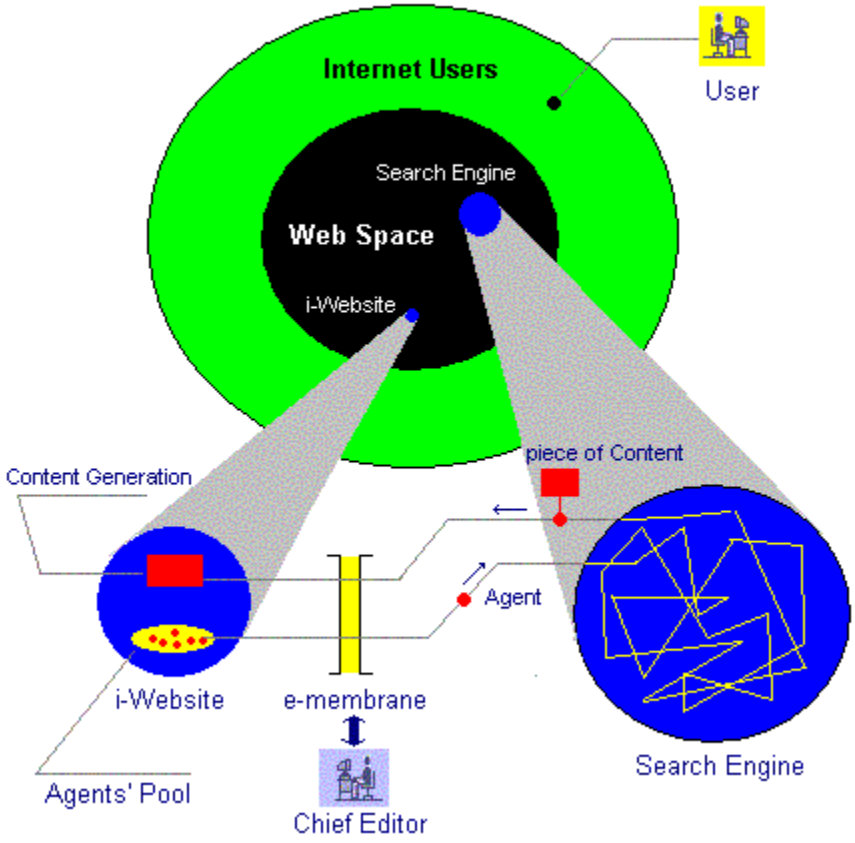
In the figure below we may schematically see how an e-membrane works. The Cyberspace is split in two realms, the “Web space” Realm black region where all the Websites are hosted and the users’ Realm green region where users are connected to Internet. The Web space is shown black because it is like the universe space, black, immense and with myriads of dispersed celestial bodies that “light” either by themselves or by the relative importance (popularity) they assign each other. Search Engines index these quasi celestial bodies and rank them following different criteria but more or less related to the “light” they emit. I-Website stands for “intelligent Website” because endowed with an Expert System, a family of agents and an e-membrane behaves much like a living Website.




The Expert System, the agents’ family and the e-membrane are the three distinctive pillars of FIRST, which stands for Full Information Retrieval Systems Thesaurus. A given Search Engine and an i-Website are amplified to appreciate their interrelation. The e-membrane has an endoderm, a mesoderm and an ectoderm and its “permeability is controlled by a human, as Chief Editor, an observer with pretended equanimity. As in living beings endoderm “manages reactions”, ectoderm “identifies and senses interactions” and mesoderm “controls” man-machine communication.

Within the agents family, one of them, named “procurebot” has the responsibility to go to the Web Realm of the Cyberspace to look for specific pieces of content. It may be instructed either to browse the whole Web by itself or to locate those pieces indirectly via the search engine indexing system. The procurebot task is guided by the Curriculum of the Content we want to generate, at large a “naked” Logical Tree (LT).

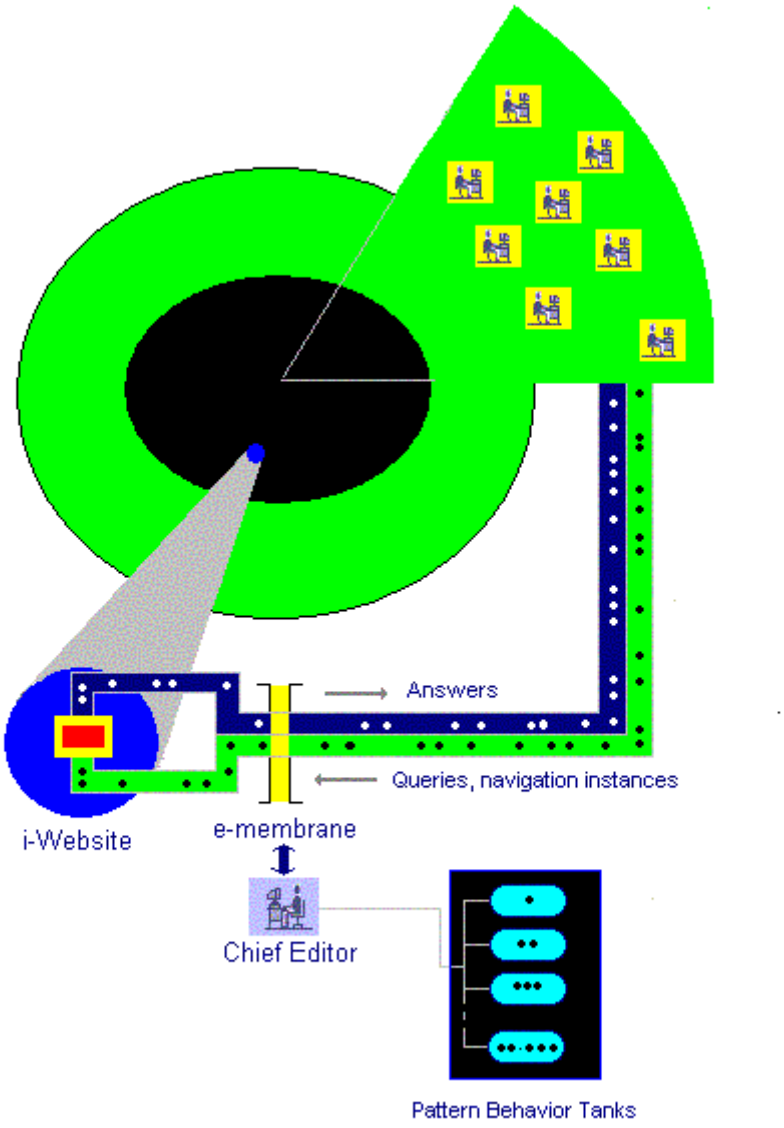
In this sense the curriculum or Index of a given discipline (LT) behaves like the DNA within the nucleus of the cell, guiding the protein buildup, and in our case guiding the cognitive content buildup. So in three steps (see “How through e-membranes Intelligence is recovered”), the Chief Editor controls and approves the agent retrieving task generating the Thesaurus (TH) and the sample of authorities for each branch and leaf of the tree (red rectangle).



# *The e-membrane in action*

## Users' Behavior Patterns Detection and Classification

Here we see the e-membrane working at its full capacity. Now the i-Website content is offered as a general attraction, a Cognitive Offer to satisfy the users' Information Demand. Users symbolically "query" the content and the content "answers". In the mesoderm region a special combinatorial algorithm generates online and in real time all possible "threads" and basic users' search strategies sequences, and stacked in Pattern Behavior Tanks. At large these patterns will resemble a sort of "Tarzan Jargon" at the users' side, a way to play a learning game against the "Oracle" who pretends to know almost everything.



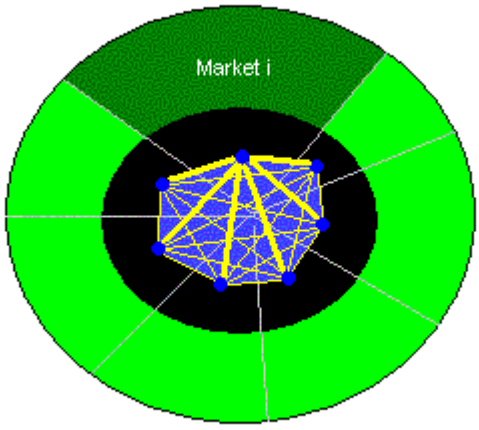

The most frequent basic search strategies sequences have a high probability to be behavior patterns. We use here human and agents talents at their best performances: agents are good to detect and to classify basic search strategies and humans are good to recognize if a given sequence could be a behavior pattern. On the other side humans are very inefficient even to imagine others basic searching strategies.

Any human plays well reacting against machine answering, from question to question, but he/she is generally poor imagining others sequences without effectively playing. Agents compute threads by billions and from time to time suggest to the Chief Editor the most probable behavior patterns.

Within the mesoderm users' tracks are split in keywords, pseudo keywords and navigation instances. Pseudo keywords are links selected by users and navigation instances are all imaginable but identifiable user actions out of querying the system, like for instance saving and downloading files, waiting specific time intervals, printing something, carrying content to a shopping cart, asking for "more" when presented answer briefs, making click over an offered document link, exits and returns to sections, login and logoff, etc. Special internal agents proceed to change the layout of the pages content in order to differentiate the users' eyes and "mood" preferences.

At large i-Website administrators become to learn from users and at the same time start to learn as much as possible (avoiding intrusive strategies) about their people behavior patterns and they could reinforce through a positive feedback a win-win scenario with their users, for instance inviting them to join autonomous Groups of Interest and Affinities of users that share similar behavior patterns. As it is depicted in the figure what a given i-Website "learns" depends of its market. It means that the same content may evolve differently when serving different markets. The rectangular yellow region over the red content corresponds to the part of content that emerged as a result of learning and evolution.

## *Darwin Nets of Learning*

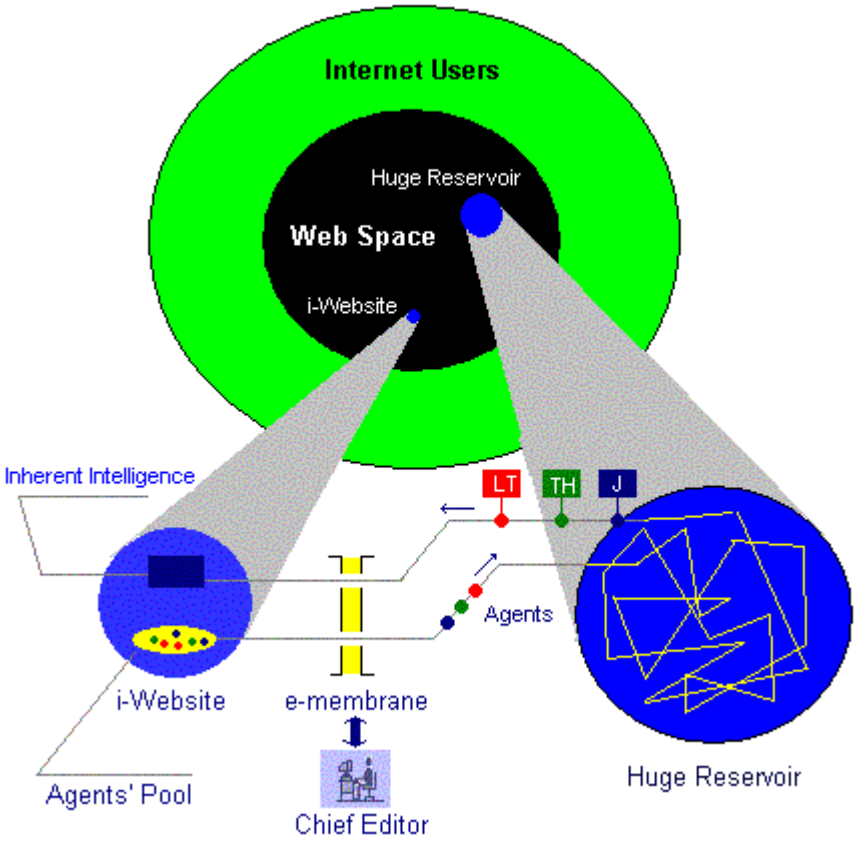



This figure represents a Darwin network which stands for Distributed Agents to Retrieve the Web Intelligence, in fact a network of i-Webs. For this reason our technology is titled Darwin-FIRST. Each node adapts to its market via customized procurebots and coopbots, other members of FIRST agents' family, take care of cooperation tasks among nodes and of the whole market intelligence equalization. Coopbots feed nodes with markets differences under the command of a Super Chief Editor.

# *How through e-membranes Intelligence is recovered*

The architecture and the content structure of huge Data Reservoirs frequently changes along the time. The old experiences are in the same reservoir of the recent ones. However the reservoir's "Inherent Intelligence" is up there and probably never changed. It could be defined as the triad Jargon, Thesaurus and Logical Tree. Most documents hosted in the reservoir are written in a given specialized Jargon, using concepts belonging to a Thesaurus and dealing with themes, topics and subjects belonging to a Logical Tree. This triad is normally dispersed throughout the reservoir and has to be recovered. With this triad everything could be indexed in an optimum way and be directly accessed. If eventually the reservoir were destroyed or spoiled the inherent intelligence enables the continuity of the tasks with the same "spirit" and under the same policies. Once this intelligence is recovered the present and future tasks will be greatly facilitated because for each subject the antecedents of it are offered in a complete and ordered way. Ideally the database architecture should adapt as close as possible to this intelligence pattern.

In the figure above three different agents backup by three different algorithms and three sets of utilities perform the recovery task. The Jargon (J) is identified and precisely defined first, then the Thesaurus (TH), and finally the Logical Tree (LT). The LT should match and cover the products and services provided by the company/institution that owns the reservoir at its deepest level of detail.



## *How Agents changes the concept we have about "Computability"*

How humans try to solve problems via computing and network facilities

Any problem could be tackled at a reasonable subjective "best" level using computing and network facilities by layers and at stepwise mode within each layer, being a step either an estimation/guess or an exact/analytical solution computation, and a layer an autonomous sequence of elementary processes of estimations and exact computations. Layers interact between them either synchronically or a-synchronically. Society as a whole accepts these reasonable good results to keep its working continuity.

With this reasoning the concept of "computability" should be reviewed. Organizations solve their problems at hand in despite of what experts think about computability. With this criterion any agreed/accepted "guess" could be assimilated as "computable" with all imaginable reserves. What is really important is that these human "intelligent" procedures could be emulated by agents and in some cases improved. Once in agents' hands organization problem solving procedures evolution could be accelerated by e-learning.

Humans are weak when performing certain open "creative" tasks such as imaging people's behavior patterns, for instance players' moves in strategic games, because of too much bound to the "reasonability" of them. Humans worry too much about their soundness and righteousness to dare imaging rare reasoning paths going from light to heavy and fuzzy ones. Agents could be settled and tuned up to ignore those "prejudices". They could be taught to explore the uncertainty space along all its dimensions without any restrictions along all possible paths, systematically, precisely and fast. Humans, on the contrary, are good for judgments. As in complex strategic games like chess humans may judge in an instant if a given "position" is either good or bad for whites or blacks, being this task extremely difficult and imprecise for agents to perform.