

The Web as it is today

The Darwin paradox: How to become smart with Not So Smart agents

Dr. [Juan Chamero](#), Madrid, Spain 27th of May 2008; Reviewed at Buenos Aires, Argentina 22th of February 2009

The Web as today is not semantically indexed at all. Even the best general Search Engines like Google, Yahoo, Altavista, etc., only index by words. Their robots inspect all pages word by word but ignore how to unveil Web page topics and how to detect “meanings” inside them: so meanings are invisible for actual search engines robots. Some Search Engines offer a limited human readable Directory of Summaries edited by human volunteer’s networks like for example Open Directory (DMOZ). There were some other works along the line of “conceptual search” -many of them defunct- like for instance Vivisimo in scientific matters and Technorati now oriented to index the growing “blogosphere”.

Many important general Search Engines failed when tried to index pages thematically. The main reason is the complexity of human mind. Any “good reader” and of course any librarian has talent enough to write an understandable summary of any book in half an hour or less in the average. The same happens with pages on the Web. However this intelligent task has not been successfully performed by agents yet. And as the Web reservoir is so huge and in continuous expansion “semantic indexing” of pages should be automatically performed by agents ideally at the moment of their upload/update. Considering that active Web pages are now in the order of 10,000,000,000 indexing via humans would mean 5,000,000,000 hours!. Thematic indexing via agents would be then the unique massive solution but the problem is that agents are not smart enough yet!. If they were smart enough a page summary could be computed in a few microseconds in the average. Hypothetically being this agent “state of the art” attainable the problem would be under control: the whole Web could be semantically indexed (by meanings) in a few hours instead!.

Darwin solved this paradox with the following stratagem: transforming apparent chaos in an apparent order!. In the Kinetic Theory of Gases individual particles within the cylinder of a gas engine behave chaotically but seen as a whole pressure vectors are working and virtually “seen”. NTS, Not Too Smart agents settled, fitted and trained under a given ontology could review thousands of pages suspected as similar and extract statistical patterns.

The ontology teaches them how a reasonable well written content would look like. As they are absolutely obedient but NTS they accept that suspected pages were all “well written” as per the master ontology. However as in the case of kinetic theory the “well_written_ness” is probabilistically distributed. In a first step Darwin used anthropic algorithms where NTS agents were assisted in smarter tasks by humans and in a second step the specific human talent was transferred to agents becoming ALS, A Little Smarter. The human “smarter” contribution was to be acquainted of coherence of conjectures. Based for example on Conjecture 1 NTS agents consider that a given cluster of pages is semantically homogeneous and a new family of also NTS agents may performs a sort of data mining on suspected meanings found within each member of the cluster. Then now either a human or a ALS agent may detect that Conjecture 2 was not well accomplished and accordingly may draw as a conclusion that the cluster is not enough homogeneous in order to unveil from it meanings. Once a design level of homogeneity is obtained, Darwin Ontology guides agents to virtually “highlight” all potential concepts as if they were human experts. This task is a huge and necessary “Knowledge Mapping” that up to the moment has never been attained.

Some leading organizations like W3C have developed tools, languages and standards that will enable Internet users to manage the whole Web as theirs. However something crucial is yet missing: the ordering of the actual Web as_it_is today, imperfectly human!.

Next reading advised

[Darwin Guide](#), something of history and Darwin Overview

[Semantic Search](#), Philosophical approach - From ideas to concepts

[Darwin versus W3C](#): two polar but complementary strategies