

Concept Processing Engine

PLATFORMedia has developed a proprietary application that enables computer users to access documents stored in their corporate LAN, indexed by 'concept'.

For instance, if a corporate financier wished to find all documents on his department's system related to 'distributed computing', our application would find word processor files, emails, presentations, etc., that contained paragraphs relating to 'distributed computing' - even if they referred to this concept as 'parallel processing', or by the name of a specific company.

Our application uses advanced statistical inference, Bayesian and other proprietary artificial intelligence and optimization techniques to enable each PC on a computer network to build a 'mental map' of the concepts that its user is interested in. To perform this complex analysis, our application harnesses the 'free' power of idle computers on network through a 'peer-to-peer' power sharing technology (also developed by PLATFORMedia). Thus, a network of computers continuously analyzes the data it contains, using each computers' excess capacity.

I : Features

PLATFORMedia's Search by Concept application is a software-only system that :

- Can extract the content of a wide variety of document types (Word documents, email, etc.);
- Refines its understanding of the documents available during a PC's "idle times", learning its user's preferences;
- Communicates with other PCs on the network to leverage the power of the whole corporate network : Its processing power increases with the number of networked PCs, and its accuracy grows as users on the network continue to make queries;
- Requires no specialist installation, and is platform- and language-independent.

II : System Requirements

This application:

- Requires a Pentium-class computer, with >20Mb free HD, 32Mb RAM
- Is written in cross-platform Java 2.0 (Windows version released first), using Sun's Java Runtime Environment 1.3 (available free)
- Includes interface code to Microsoft Office applications written in C++
- Runs as a background process on each PC, to enable off-line processing

- Interfaces with a standard SQL database, notably the open-source MySQL (available free)
- Implements a 'client' side HTTP server on each PC that it runs on, so that it can be accessed through a standard Browser
- Is built on top of PLATFORMedia's Peer-to-Peer infrastructure, which enables distributed processing of tasks within the corporate LAN – with each PC communicating in XML

III : Intellectual Property

PLATFORMedia has developed the technology in-house, and owns the following fundamental technology :

- The mechanism to derive plain-text results from a wide variety of different input sources :
 - Microsoft Word (v2.0, 95, 97 and 2000), RTF and TXT documents
 - Internet/Intranet documents in HTML
 - Microsoft Outlook emails and attached documents
- The ability to pre-process these files, to derive meaningful 'units' of data (typically paragraphs)
- Statistical analysis techniques that can determine the relevance of collections of words to each other, and with other collections of words
- Optimization techniques that dynamically reorganize the databases to save disk space and reduce access times
- Evolutionary optimization techniques to address the problem of finding significant 'concept clusters' from the databases
- Collaborative processing via PLATFORMedia's Peer-to-Peer infrastructure with other PCs in the same organization

IV : Applications

The concept search system can be used in a variety of applications:

- Search by individual users across their corporate knowledge-base;
- Automatic categorization of documents into subject areas;
- Automatic individual daily news pages, filled with relevant Internet links;
- Discovery of overlapping individual expertise within large corporations;
- Content-to-user matching systems on Internet sites;