

Personalized Information System (PIS) Patent Filing

Title:

System and Method for Personalization of Information (Abstracts/Articles)

Description

Field of the invention:

The invention, PIS is an application that is deployed over web. PIS can be implemented as a research tool that can be use for the biomedical community, petrol-chemical industry, business intelligence and political risk assessment.

The objective of the information is that the users can retrieve information without using a search engine but instead have information route to users based on the users' profile.

Background of the invention:

The ("Users") may include researchers in their respective fields, librarians, business advisors, consultants in their respective fields, students, lecturers.

The users can access the invention ("PIS – Personalised Information System") via web from home, work or school.

The ("Web") refers to the World Wide Web; a client program called a web browser retrieves information resources, such as web pages and other computer files, from web servers using their URLs and displays them, typically on a computer monitor. One can then follow hyperlinks in each page to other resources on the **World Wide Web** whose location is provided by these hyperlinks. Web pages are often arranged in collections of related material called "websites." [Wikipedia]

The invention consists of taxonomy, text categorizer and user profiler.

A ("Taxonomy") is a hierarchical set of words that either have a parent to child relationship (as found in Set Theory) or may be related due to some other forms of relationship defined by the person creating the taxonomy. Due to the relationship between the words, taxonomy (singular) or taxonomies (plural) are often used to represent knowledge of a certain subject or faculty ontologically. This knowledge representation is normally used for categorization or classification purposes. See Fig A-1.

The ("Application Taxonomy") refers to the taxonomy that is created for a specific application for the various industries such as, the bio-medical, oil and gas discovery, and finance related etc.

A (“Text Categorizer”) is a set of codes that are programmed to understand a (“Document”) electronic document’s topic. The electronic document format has various file extensions such as .txt, .doc, .pdf, .emf etc. A text categorization (also known as document classification) tasks can be divided into two sorts: supervised text categorization where some external mechanism (such as human feedback) provides information on the correct classification for documents, and unsupervised text categorization, where the classification must be done entirely without reference to external information.

A (“Document”) includes research documents, research papers, articles, abstracts, in the various formats.

A (“User Document”) belongs to a specific user (associated to a user) and related to the user area of research and area of interest. It may include research documents, research papers, articles, abstracts, in the various formats.

A (“User Profiler”) is a set of codes that are programmed to understand a user’s area of interest or the user’s area of research. The user profile will help to match words (categories) in the taxonomy to be associated to a user.

A (“User Profile”) will be determined by the user’s personal information and the user’s documents that the user uploads into the invention’s database (or file system).

A user will usually experience getting huge amount of search results whilst searching for certain information. The huge amounts of search results are due to a few factors. One is because of the wrongful categorization electronic documents and web pages.

Hence when a user keys in “Eye” in a search engine, the user can get results related to the Eye – the anatomy, Eye – the private eye or Eye – the eyewitness and the list goes on.

A word search can also be erroneous because of the words used that were keyed in the search engine. A user may have limited knowledge of the word that the user is searching for. Hence the user may key in the search engine a related but not accurate word when using the search engine.

Combination of these 2 factors will eventually include many unrelated electronic documents and web pages being listed amongst the search results.

The invention instead “guides” the user to the information rather than the user himself/herself searching for the results.

The invention firstly categorizes the documents (abstracts/articles) which are in the form of electric document into appropriate categories through a text categorizer.

The text categorizer is trained by using training sets (or exemplary electronic documents) for a particular category. The text categorizer extracts words from these training sets and

use it as a benchmark as to whether accept or reject a new document to be under a category that is under the taxonomy.

A collection of documents are then uploaded onto the invention's application database (or file system) and are then indexed. These documents can come from the web, documents from the user's local drive or documents that have been saved in a database (or file system). Documents are then categorized based on the taxonomy that can be developed by a domain expert or taxonomy that are created for the general public use. For every document uploaded into the database (or file system), there will be a category associated to the document.

The user profiler uses the text categorizer to determine the categories that are related to a user. A user may have one or more categories associated to the user.

The working method in the user profiler is as follows:-

Firstly, the user would have to key certain personal information when registering to use the invention. Part of this personal information includes the user's area of research and area of interest.

The invention will save the texts that are being keyed in by the user as a personal information document. Next, the invention allows the user to upload user's documents that are related to the user's work or research scope. Both the text of the user's personal information and the documents that the user uploads goes to the inventions database (or file system).

The aggregation of the user's personal information and the user's documents that the user uploads will then be categorized using the same taxonomy that was used to determine the collection of documents' categories.

The categorization of the personal information will return category/ categories with respect to the personal information documents. Likewise, every individual user's document will also return category/categories with respect to the user's document.

Categories returned from the personal information document and the user's documents will be accumulated into a list of categories. The list will check for duplication of categories within the list and will remove any duplication.

The remainder of the list will be checked with categories in the application taxonomy and if the categories in the list created is not available in the categories found in the application taxonomy, then the categories in the list will be deleted.

The remainder of categories in this list forms the user's profile. These categories are then matched to categories found in the application taxonomy.

Under the categories in the application taxonomy are documents that are associated with each category.

If a user has categories under his profile, then these categories under the user profile will match to the categories found in the application taxonomy which in turn is matched to documents that have been categorized under the same category.

The relationship is as follows:-

Category A in User Profile → Category A in Application Taxonomy → Documents
under Category A in Application Taxonomy