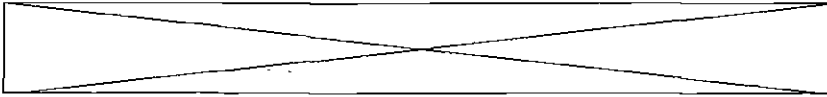


# EXHIBIT C

**From:** Saved by Windows Internet Explorer 7  
**Sent:** Thursday, November 20, 2008 3:58 PM  
**To:**  
**Subject:** LexiClone.com | Patent #6.199.067, The Technology Overview  
**Attach:** ATT00692.gif;ATT00695.gif;ATT00698.gif;ATT00701.gif;ATT00704.gif;ATT00707.css

---



[Home](#) | [News&Press-releases](#) | [Products](#) | [Company Overview](#) | [Investor Relations](#) | [Order&Buy](#)

[Our Philosophy](#)  
[Technology](#)  
[Articles](#)  
[Our Team](#)  
[Contact Information](#)

## The Patent Technology Overview

### What the Patent #6.199.067 covers:

- The core summarizing technology
- Data searches using the core summarizing technology performed between two computers
- Data searches using the core summarizing technology performed between a computer and the Internet
- Data searches using the core summarizing technology performed on a single computer.

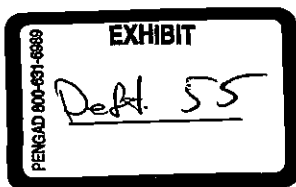
### Basic Background of the Field of the Patent: Why is the Technology Needed

In recent years, computers have taken the world by storm. Today, most businesses entirely rely on computers to conduct daily operations. In the academic world, computers have become essential tools for learning, teaching and research. In homes, computers are used to perform daily tasks ranging from paying bills to playing games. The one unifying requirement for all computer applications is the ability of a user to utilize a computer to locate particular information or data desired by the user. Before LexiClone the only known way to find information is to search for key words.

A number of approaches have been developed to improve the performance and accuracy of typical key word searches. For example, U.S. Patent Number 5,845,278, issued to Kirsch, et. al, teaches approaches to establishing a quantitative basis for selecting client database sets (i.e. Internet documents or web sites) that include the use of comprehensive indexing strategies, ranking systems based on training queries, expert systems using rule-based deduction methodologies, and inference networks. These approaches were used to examine knowledge base descriptions of client document collections or databases.

However, the key word searching approaches utilized by previously known search engines suffer from a number of significant disadvantages. Most search systems are viewed as often ineffective in identifying the likely most relevant documents. Accordingly, the users are often presented with overwhelming amounts of information in response to their key words. Thus, using proper key word searching techniques becomes an art in itself - an art that is outside the capabilities of most Internet users.

Most importantly, typical key word and even more advanced searches only provide the user with search results that depend entirely on the search string entered by the user, without any regard to the user's cultural, educational, social backgrounds or the user's psychological profiles. The results returned by the search engines are tailored only to the search string provided by the user and not to the user's background. None of the previously known search engines tailor results of user's searches based on his or her background and





unexpressed interests. For example, a twelve year old child using key word searches on the Internet for some information on computers may be presented with a multitude of documents that are far above the child's reading and educational level. In another example, a physician searching the Internet for information on a particular disease may be presented with dozens of web sites that contain very generic information, while the physician's "unexpressed" interest was to find web sites about the disease that are on his educational and professional level.

It would thus be desirable to provide a system and method for extracting and using linguistic patterns of textual data to assist a user in locating requested data that, in addition to matching the user's specific request, also corresponds to the user's professional, cultural, educational, and social backgrounds as well as to the user's psychological profile and thus addresses the user's "unexpressed" requests.

### **Summary of the Patent**

This invention relates to use of linguistic patterns of documents to assist a user in locating requested data that, in addition to matching the user's specific request, also corresponds to the user's cultural, educational, professional, and social backgrounds as well as to the user's psychological profile, and thus addresses the user's "unexpressed" requests. The present invention provides a system and method for automatically generating a personalized user document's summary of "key" phrases based on linguistic patterns of documents provided by the user and for utilizing the generated summary to perform adaptive Internet or computer data searches.

The system of the present invention advantageously overcomes the drawbacks of previously known data searching techniques. As was noted earlier, typical key word and even more advanced searches only provide the user with search results that depend entirely on the search string entered by the user, without any regard to the user's cultural, educational, professional, and social backgrounds or the to user's psychological profile.

All texts composed by the user, or adopted by the user as favorite or inimical (such as a favorite book or short story), contain certain recurring linguistic patterns, or combinations of various parts of speech (nouns, verbs, adjectives, etc.) in sentences that reflect the user's cultural, educational, social backgrounds and the user's psychological profile. Research has shown that most people have readily identifiable linguistic patterns in their expression and that people with similar cultural, educational, and social backgrounds will have similar linguistic patterns. Furthermore, research has shown that such factors as psychological profile; life experience, profession, socioeconomic status, educational background, etc. contribute to determining the frequency of occurrences of particular linguistic patterns within the user's written expression.

In accordance with the present invention, particular linguistic patterns and their frequencies of occurrence are extracted from the texts provided by a user of the system of the present invention and stored in a user summary data file. The user data file is thus representative of the user's overall linguistic patterns and their respective frequencies. All documents in a remote computer system, such as the Internet, are likewise analyzed and their linguistic patterns and frequencies thereof also extracted and stored in corresponding document summaries. When a search for particular data is initiated by the user, linguistic patterns are also extracted from a search string provided by the user into a search summary. The user summary is then cross matched with the search summary and the document summary to determine whether any linguistic patterns match in all three summaries and to determine the magnitude of the match based on summation of relative frequencies of matching patterns in the user summary and the document summary. The documents with document summary having the highest matching magnitudes are presented to the user as not only matching the subject of the search string, but also as corresponding to the user's cultural, educational, and social backgrounds as well as the user's psychological profile. Thus, a world renowned physicist searching for information



on quasars would be presented with very sophisticated physics documents that are oriented to wards his level of expertise.

It should be noted that the user's background and psychological characteristics are not evident directly from the linguistic patterns themselves or from their frequencies. Accordingly, the system of the present invention matches the user's linguistic patterns to the linguistic patterns of data requested by the user without extracting any actual information about the user's background and psychological characteristics from the user profile. Thus, the user's privacy is not impinged by the creation and retention of the user summary.

The profiling/search system includes a local computer system, connected to a remote computer network (e.g. the Internet) via a telecommunication link. The local computer system includes a control unit and related circuitry for controlling the operation of the local computer system and for executing application programs, a memory for temporarily storing control program instructions and variables during the execution of application programs by the control unit; a storage memory for long term storage of data and application programs; and input devices for accepting input from the user. The local computer system further includes: output devices for providing output data to the user and a communication device for transmitting to, and receiving data from, the remote computer system via the telecommunication link. The remote computer system includes a communication gateway connected to the telecommunication link, a remote data storage system for long term data storage, and a remote computer system control unit (hereinafter RCS control unit).

In summary, the system of the present invention operates in three separate independent stages, each stage being controlled by a particular control program executed by one of the local computer system and the remote computer system. In a first stage, a user profiling control program is executed to generate or update a user summary computer file representative of the user's linguistic patterns and the frequencies with which these patterns recur in texts submitted by the user and/or automatically acquired by the inventive system. The user is then invited to provide textual data composed by the user such as e-mail messages, memorandums, essays as well as documents composed by others that the user has adopted as "favorites", such as favorite web sites, short stories, etc. These textual documents are temporarily stored in a user data file. The inventive system also monitors the user's data searching and data browsing (e.g. Internet browsing) to automatically add additional textual information to the user data file. Once the user data file attains a sufficient size, or when other criteria for updating the user summary are met, the system executes a summary extraction subroutine to create/update the user summary by extracting linguistic patterns from the user data file.

During the summary of a text extraction subroutine, the system retrieves individual textual documents from the user data file, and separates each document into sentences. The system then extracts a linguistic pattern, or a segment, from each sentence characterized by first identifying words in the sentence as being particular parts of speech (i.e. nouns, verbs, adjectives, etc.), and then selecting a predetermined combination of the identified parts of speech and storing this combination as a segment. In a preferred embodiment of the present invention, each segment comprises a triad of three parts of speech: noun - verb - adjective. The segment extraction process is repeated for all textual documents in the user data file. The system then groups identical segments together and determines their frequency of occurrence in the user text's summary. Thus, the resulting user summary contains the linguistic patterns from all texts submitted by the user (or automatically gathered by the system) and the frequencies with which those patterns recur within the texts.

In a second stage of the present invention, a data profiling control program is executed to generate data item summary computer files, representative of linguistic patterns and their respective frequencies, of all data items. The data items may include documents, web sites, and other textual data that may be subjected to a search by the user. A list of all data items and their respective data addresses (such as Internet URL addresses) is first provided to the system.



The data item summary generation procedure is then performed for each data item in the list in a similar manner to the user-profiling procedure, except that data item address information is stored in each data item's summary. Thus, the resulting data item summary of each data item contain the data item address, the linguistic patterns of the data item and the frequencies with which those patterns recur therein.

In a third stage of the present invention, the system executes a data searching program that enables a user to utilize the system to perform advanced searches for desired data files, such that the data files returned as search results correspond to the user's social, educational, and cultural backgrounds and to the user's psychological profile. The search program is initiated when the user provides a search string representative of data requested by the user to the system. The system then creates a search summary representative of linguistic patterns in the search string in a similar manner to the user-profiling procedure, except that frequencies of recurring segments are not recorded in the search summary. Optionally, the system expands the search summary by generating additional segments that contain synonyms of the parts of speech in the existing segments already in the search summary, and storing the additional segments therein.

After the search summary is complete, the system retrieves the user summary of the user performing the search and compares the segments stored in the user document summary with the segments stored in the search template to determine a number of matches between various segments in each of the templates and then, for each matching segment records the frequency with which the matching segment recurs within the user template. The system then applies the original search string to a standard match engine to obtain a list of data item addresses that potentially match the user's search requirements and then retrieves the data item templates corresponding to the data item addresses on the list. This procedure is optional but is recommended because a direct linguistic pattern search over all data items stored on the remote computer system can be very time consuming given the modern computing and data transfer technologies.

The system then compares, for each data item template, the segments stored in the data item template with the segments stored in the search template to determine a number of matches between various segments in each of the templates and then, for each matching segment records the frequency with which the matching segment recurs within the data item template. A match value is then determined by the system for each segment in the data item template that also appears in the search template and in the user template, by adding the frequency of the segment's occurrence in the data item template to the frequency of the segment's occurrence in the user template. Finally, the system computes a final value for each data item template by adding together the match values of all matching segments in each data item. The final value is representative of the degree to which the linguistic pattern of the data item matches the linguistic pattern of the user in light of the linguistic pattern and subject matter of the search string. The data items, corresponding to data item templates having the highest final values, are then retrieved by the system. The system then presents the user with several data items having the highest final values, starting with the data item with the highest final value.

---

[E-mail](#) | [Sitemap](#) | [LexiClone](#) ❄

All software and content is copyrighted 1998-2002



