EXHIBIT A

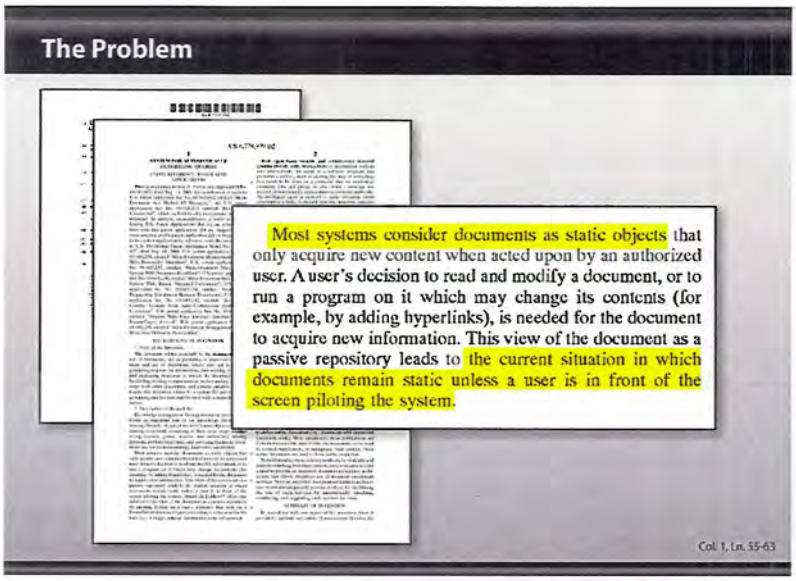
Introduction / The Problem

May it please the Court:

Defendants Google, Yahoo!, and Right Media present this technology tutorial in the patent infringement action brought by Plaintiff Xerox.

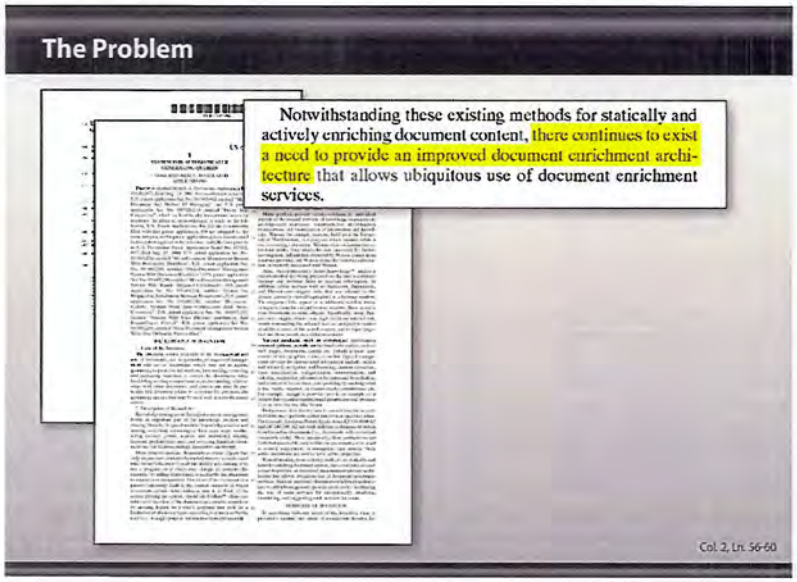
There is one patent at issue in this case: U.S. Patent Number 6,778,979.

The patent claims a method for automatically generating a query from selected document content.

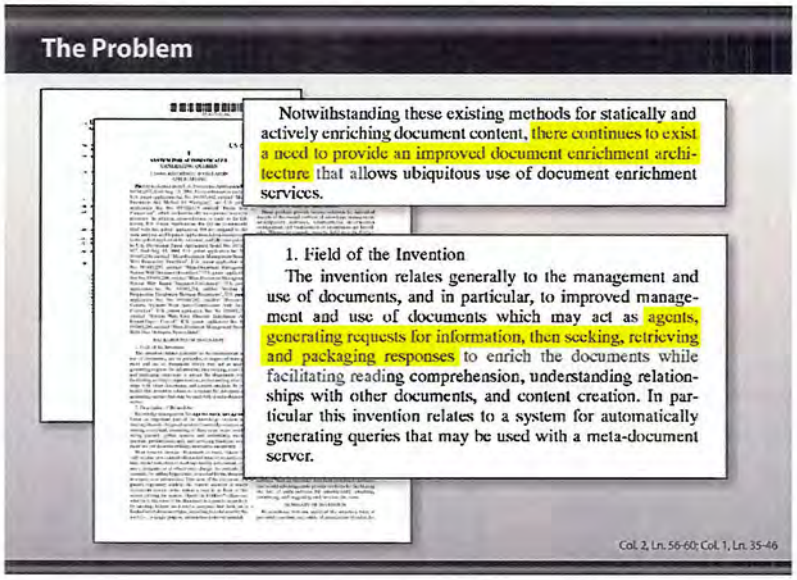


The '979 patent, issued in August 2004, relates to a system for dynamically obtaining additional information related to documents. According to the patent, most contemporary document management systems treated documents as "static objects" and "passive repositories" of information.

The Problem



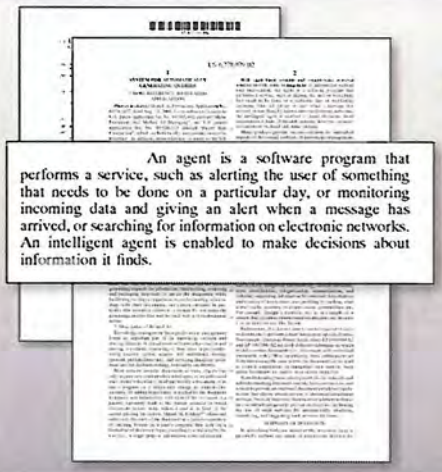
According to the patent, there existed a need for an "improved document enrichment architecture" that allowed for documents to be enriched with additional content without user intervention.



This functionality is added through the use of software agents which generate requests for additional information based on the content of the document.

Existing Solutions

Existing Solutions



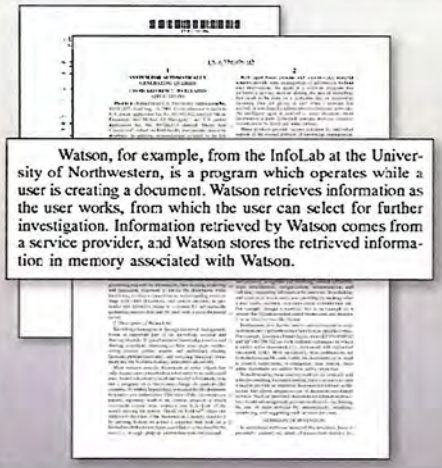
An agent is a software program that performs a service, such as alerting the user of something that needs to be done on a particular day, or monitoring incoming data and giving an alert when a message has arrived, or searching for information on electronic networks. An intelligent agent is enabled to make decisions about information it finds.

- "An agent is a software program that performs a service...."

Col. 2, Ln. 3-9

Agent-based systems are software programs that perform certain tasks based on instructions or information, acting as the user's "agents." As the specification acknowledges, prior art agent-based systems were capable of providing document enrichment services.

Existing Solutions



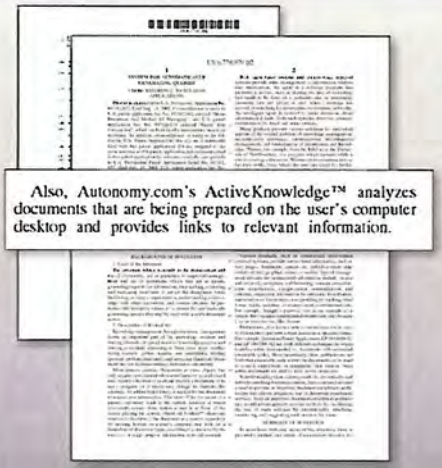
Watson, for example, from the InfoLab at the University of Northwestern, is a program which operates while a user is creating a document. Watson retrieves information as the user works, from which the user can select for further investigation. Information retrieved by Watson comes from a service provider, and Watson stores the retrieved information in memory associated with Watson.

- "An agent is a software program that performs a service...."
  - Watson (Northwestern University InfoLab) monitors the user's activity as he creates a document and retrieves relevant information from online service providers based on the user's choices

Col. 2, Ln. 14-20

For example, an agent could be programmed to search the Internet for information related to words or phrases that a user is typing into a word processor, and present that information to the user. The Watson system from Northwestern's InfoLab,

Existing Solutions



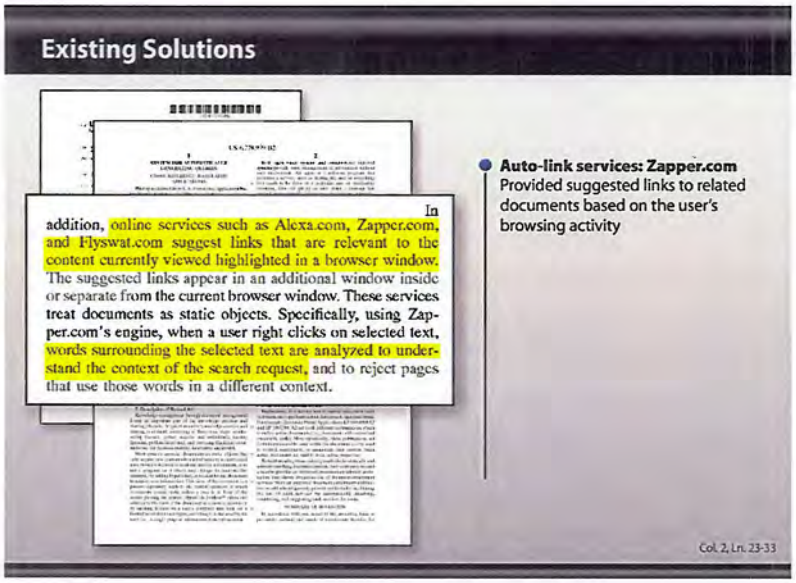
Also, Autonomy.com's ActiveKnowledge™ analyzes documents that are being prepared on the user's computer desktop and provides links to relevant information.

- "An agent is a software program that performs a service...."
  - Watson (Northwestern University InfoLab) monitors the user's activity as he creates a document and retrieves relevant information from online service providers based on the user's choices
  - Autonomy.com's ActiveKnowledge system analyzed documents created by the user to provide links to relevant information on the Web

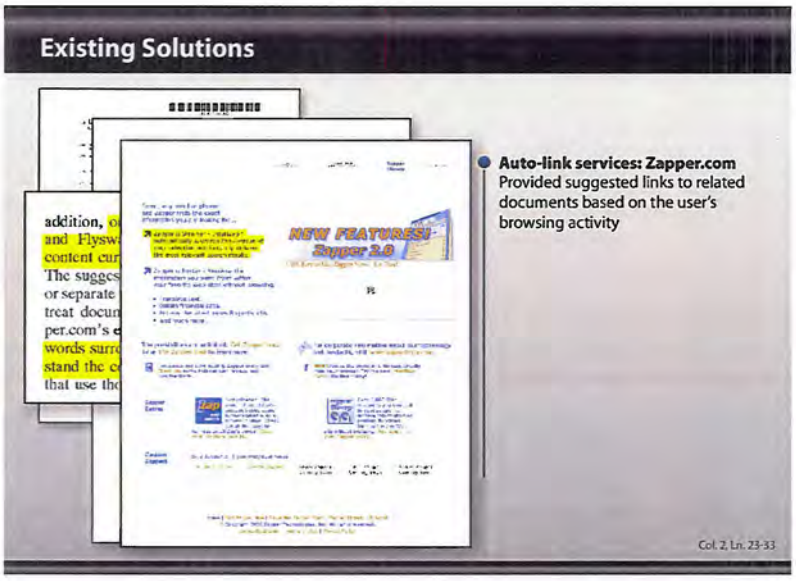
Col. 2, Ln. 21-23

and the ActiveKnowledge system from Autonomy.com are two such solutions discussed in the patent.

Existing Solutions

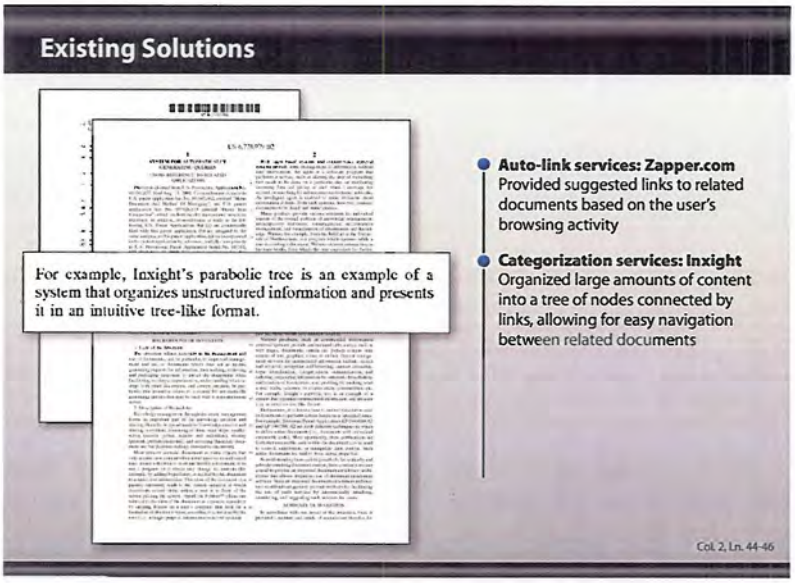


The patent also lists other existing solutions to the asserted problem of document enrichment and management. One approach employed by products such as Alexa, Zapper, and Flyswat is to scan the content of the document, locate key words or phrases, and provide additional information to the user on those phrases if requested.

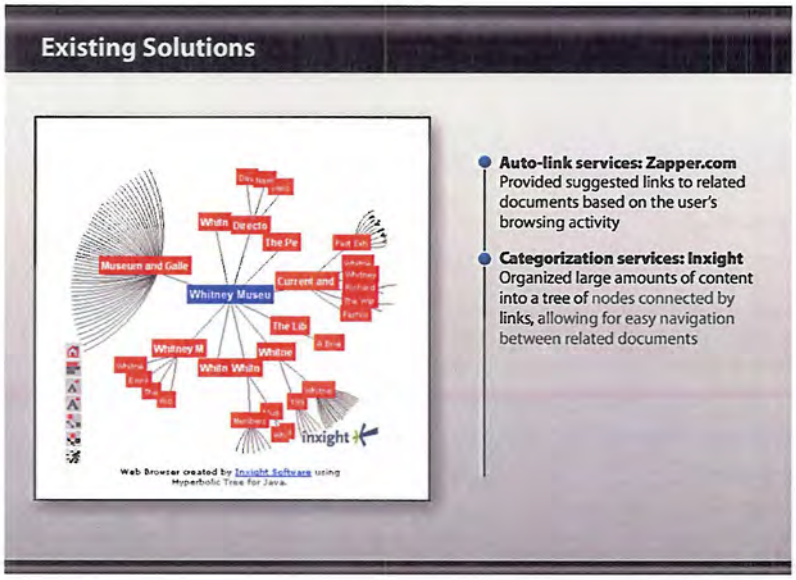


For example, Zapper might highlight text phrases like "Vikings," "Michael Jordon," or "Las Vegas" for the user. Should the user right click on the highlighted text, Zapper submits a query to a search engine for that text, then uses the words surrounding the text to filter the results.

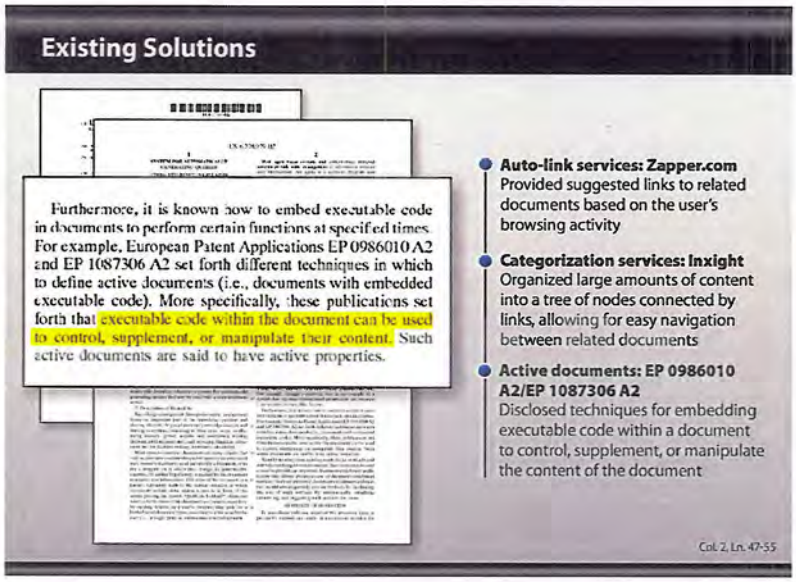
Existing Solutions



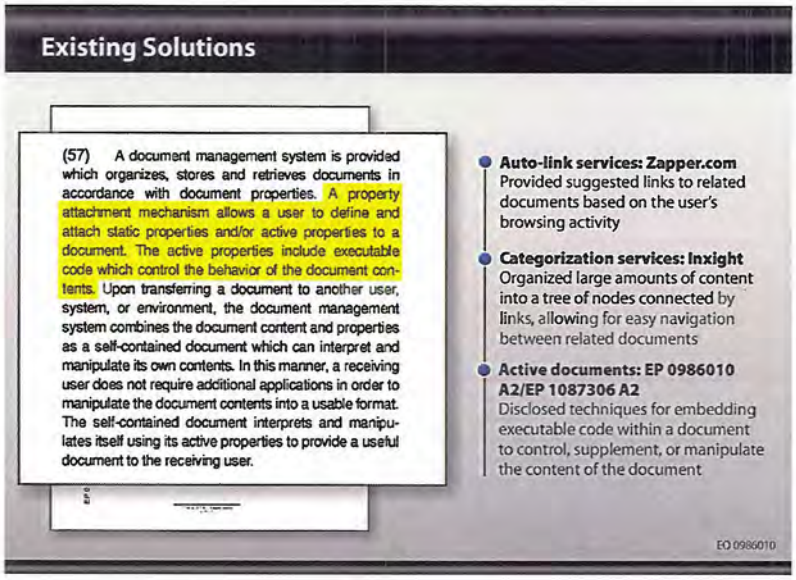
Other solutions, such Inxight's Parabolic Tree, automatically categorize the links present on a document and present them to the user in a structured hierarchy of nodes or categories.



Existing Solutions

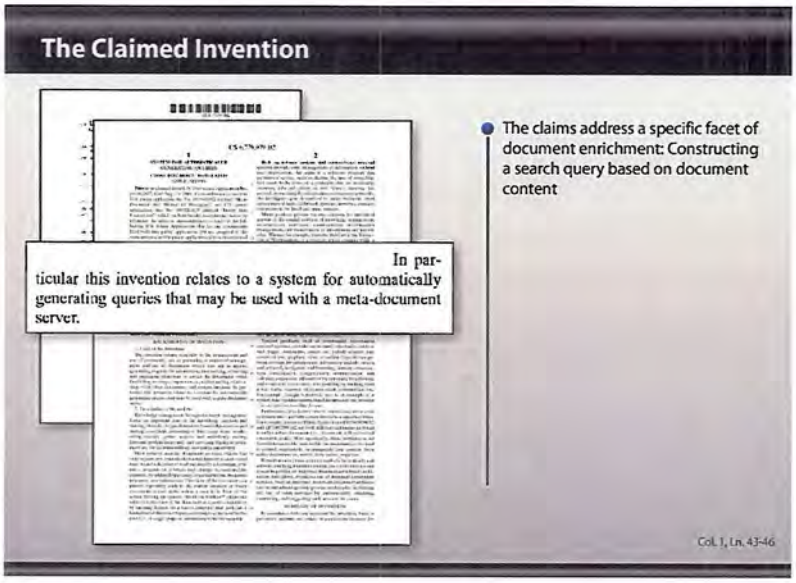


Active documents were also known in the art.

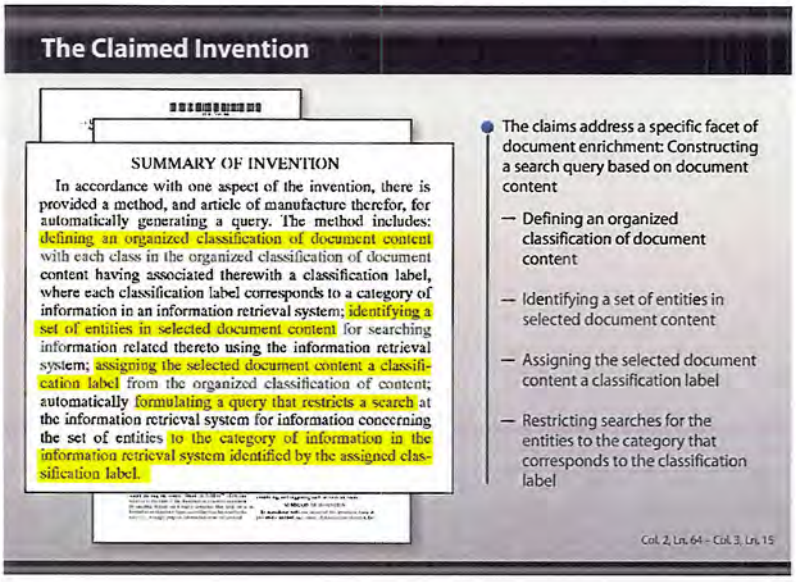


For example, the '010 and '306 European patent applications disclosed techniques for embedding executable software code within documents to manipulate the content of the documents based on a user's preferences.

The Claimed Invention

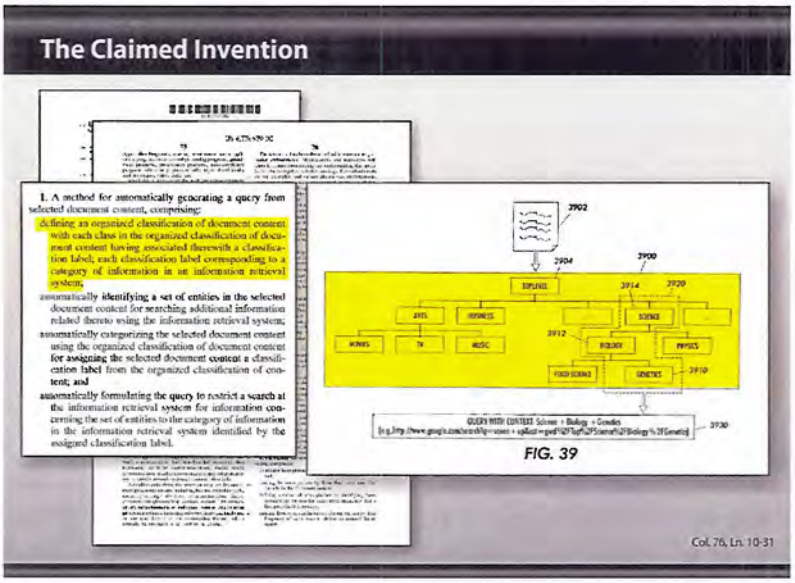


While the specification of the '979 patent describes a number of aspects to enriching document content, the claims all relate to a "system for automatically generating queries that may be used with a meta-document server" such as a search engine.

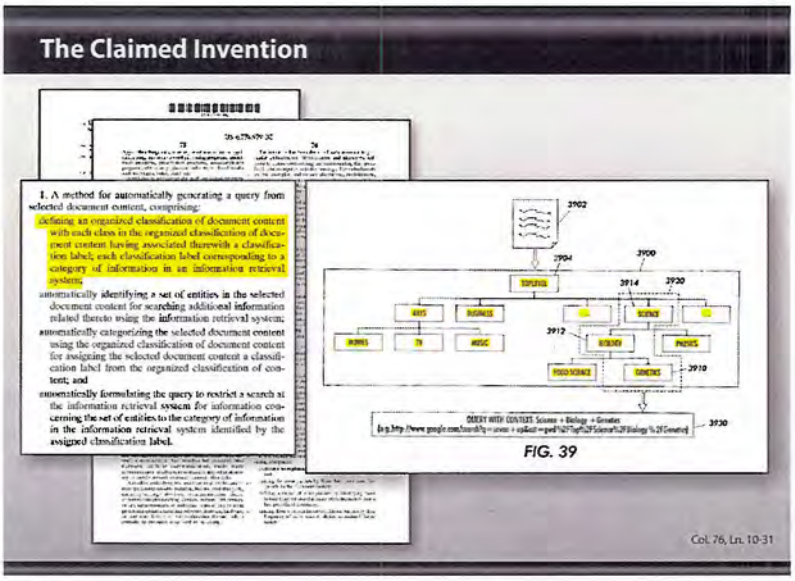


Specifically, the claims disclose a method of defining an organized classification of document content, identifying a set of entities or words within the document for performing additional searches, assigning the selected document content a classification label from the organized classification of document content, and automatically formulating a query to restrict a search at the information retrieval system concerning the identified words or entities to the corresponding category of information.

The Claimed Invention

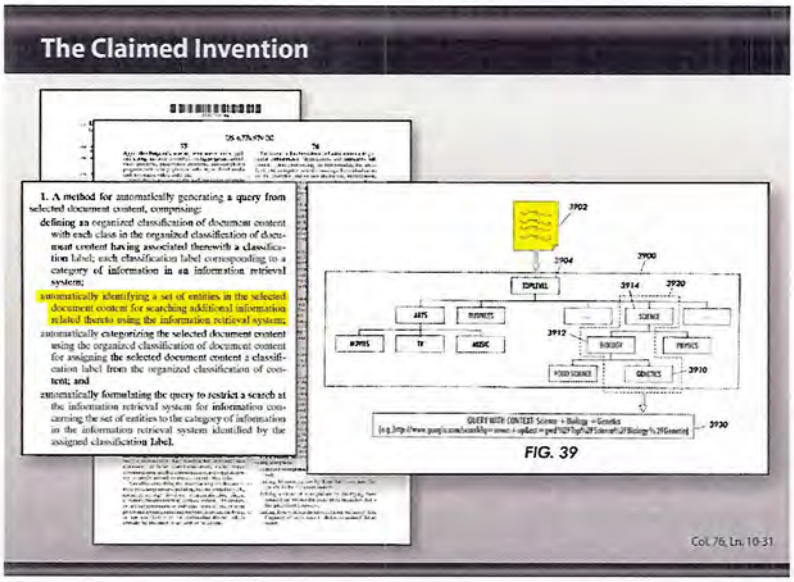


First, an organized classification of document content is defined. The classification scheme corresponds to a set of categories of information in an information retrieval system,

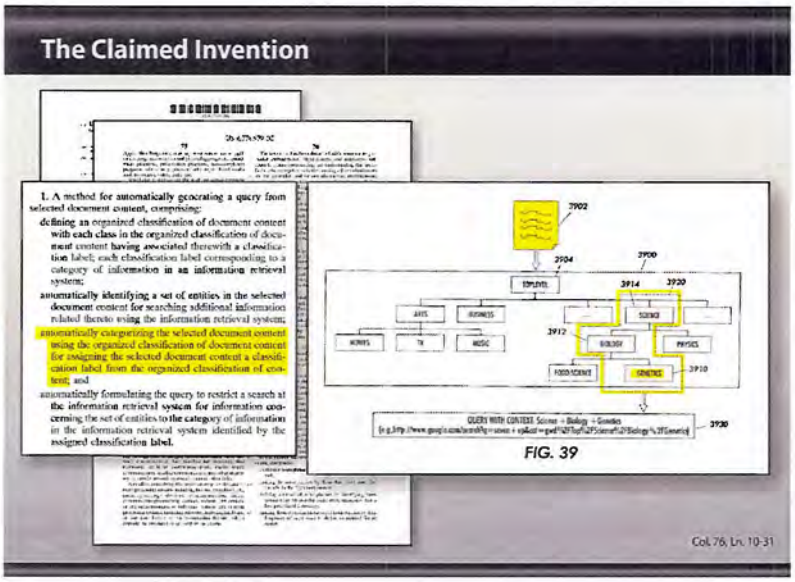


and each category has a classification label associated with it. The categorization will later be used to formulate a query to restrict a search to one assigned category.

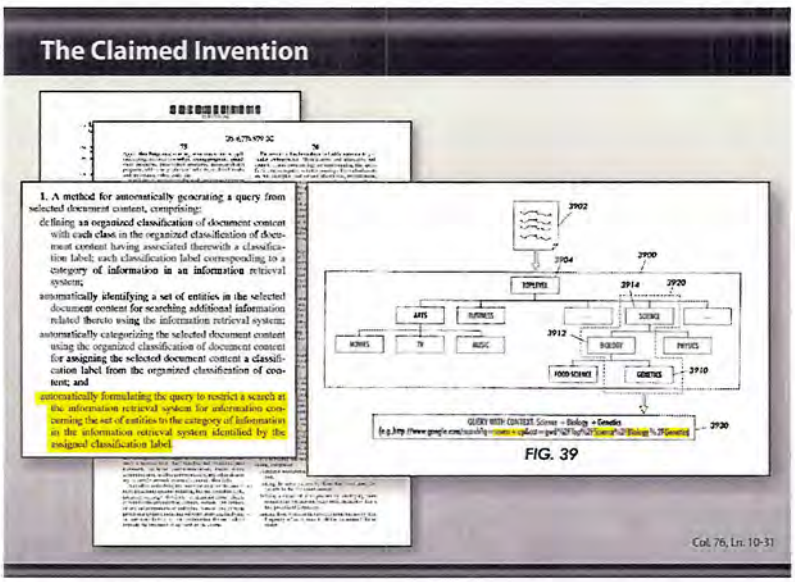
The Claimed Invention



In the second step, a set of entities – words, phrases, or other portions of the document – is identified in the content of a document. Each word or phrase represents a concept about which a user may desire additional information.

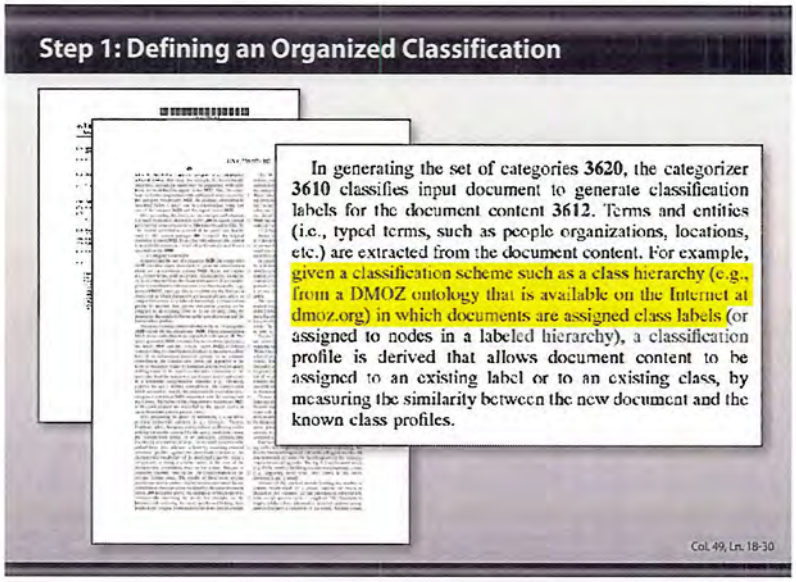


In the third step, selected document content is analyzed using the classification scheme and is assigned one classification label from the organized classification defined in the first step.

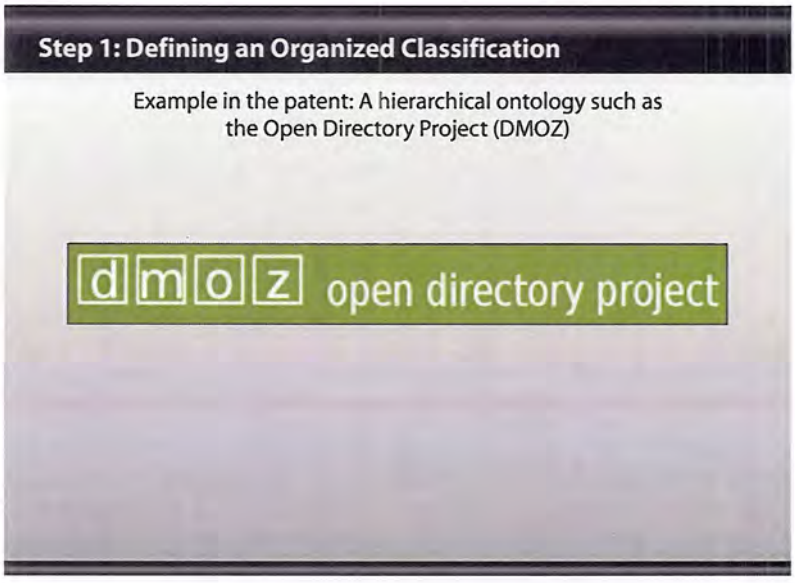


In the last step, a query is formulated to restrict a search to the category in the information retrieval system that corresponds to the selected document content's classification label.

Step 1: Defining an Organized Classification

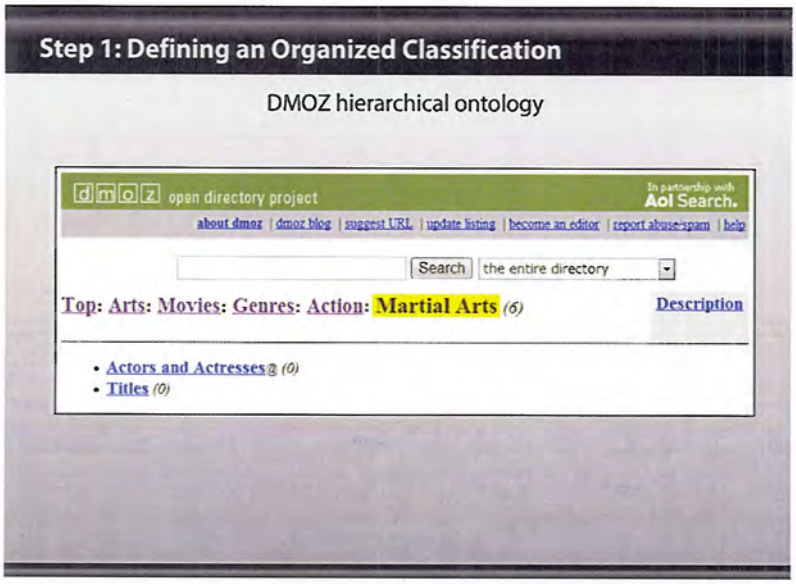


Before the system can classify document content and use the classification to restrict a search, it must first define the classification scheme to be used.

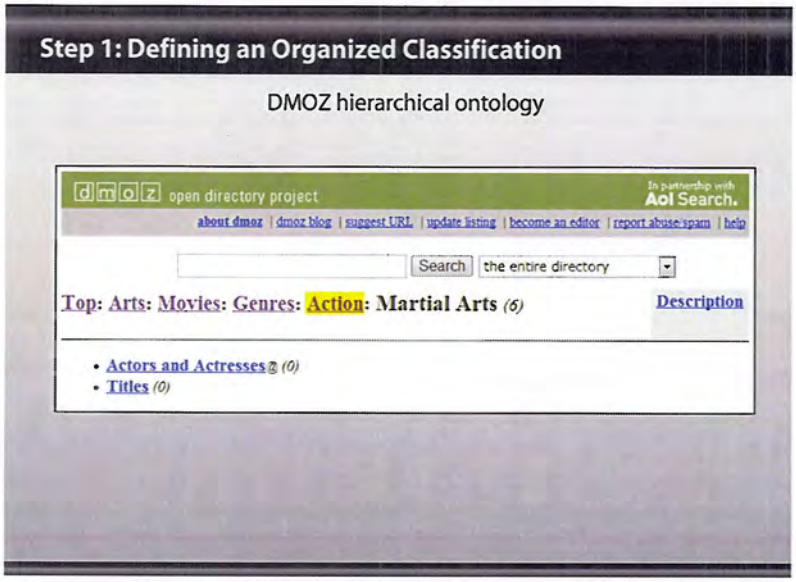


An example of such a scheme given in the patent is the Open Directory Project, a publicly available content directory of links to World Wide Web sites also known as DMOZ. DMOZ uses a hierarchical ontology for organizing site listings by subject matter.

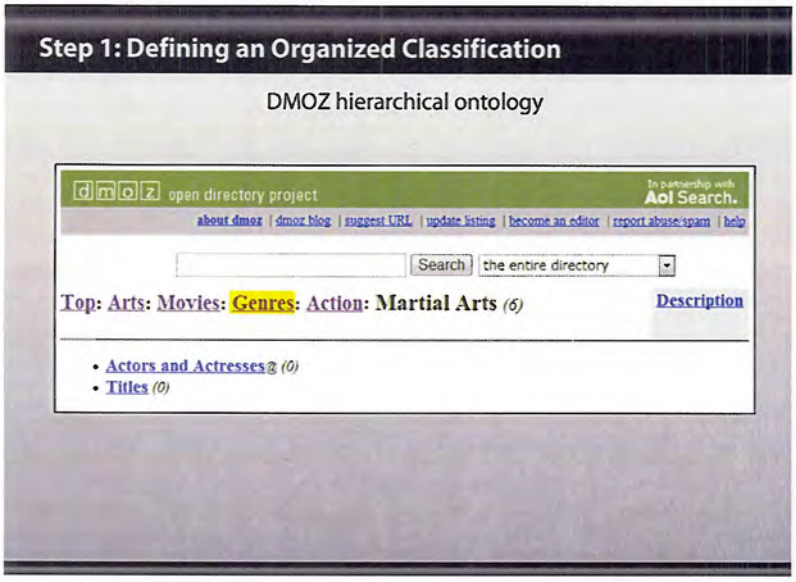
Step 1: Defining an Organized Classification



For example, a document could be classified in the category "Martial Arts," which is itself a subcategory of the category



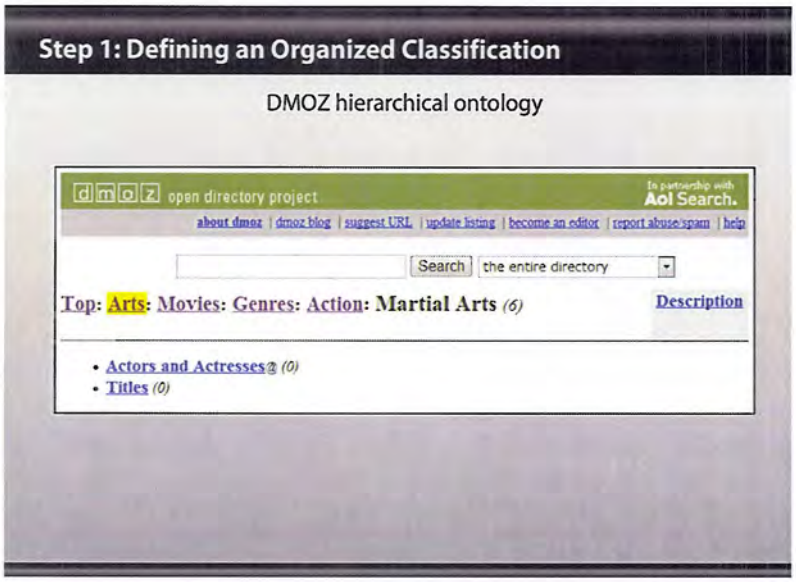
"Action," which is a subcategory of



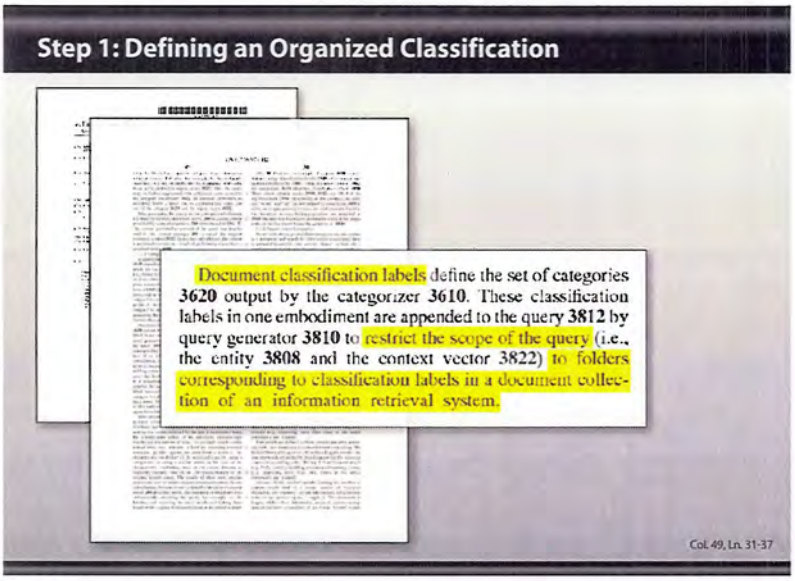
"Genres," which is itself a subcategory of

Step 1: Defining an Organized Classification

"Movies," which, finally, is a subcategory of the category

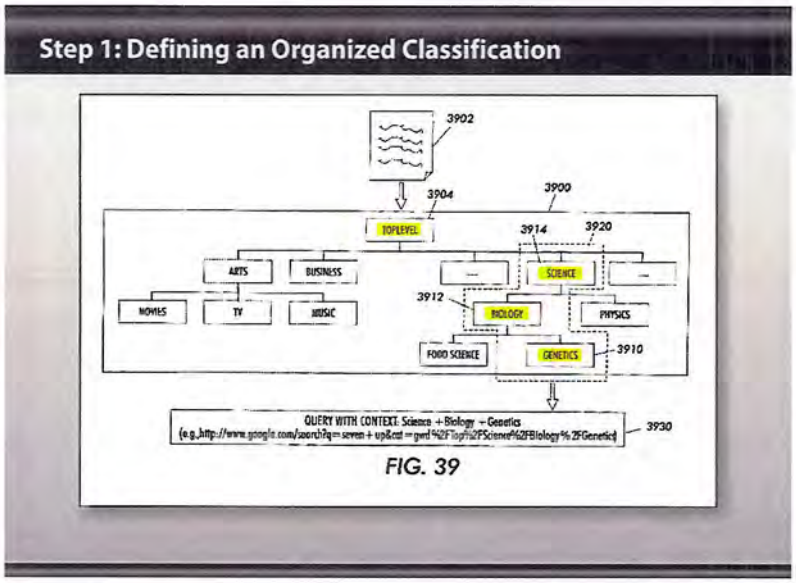


"Arts."

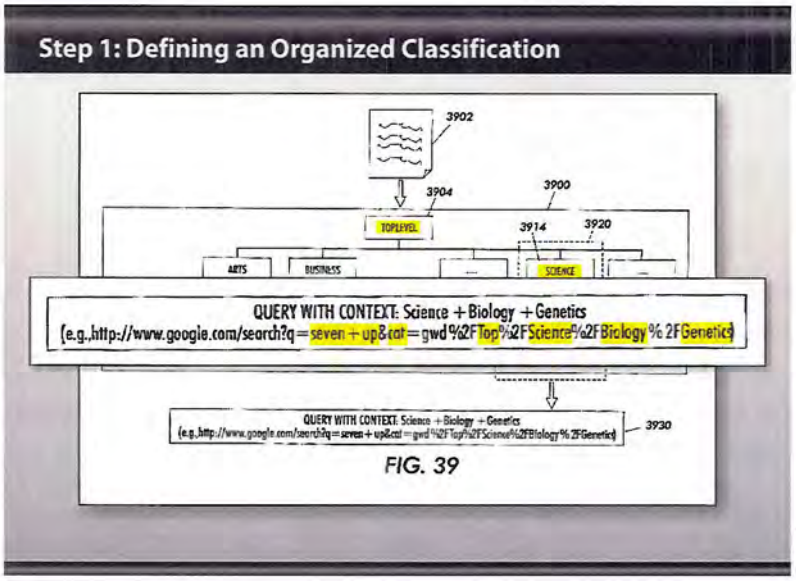


Importantly, the categories that are defined in this step for the purpose of document classification need to correspond to categories in the information retrieval system that will ultimately be searched. As we will see later on, this allows the claimed process to restrict a search at the information retrieval system to the assigned category.

Step 1: Defining an Organized Classification

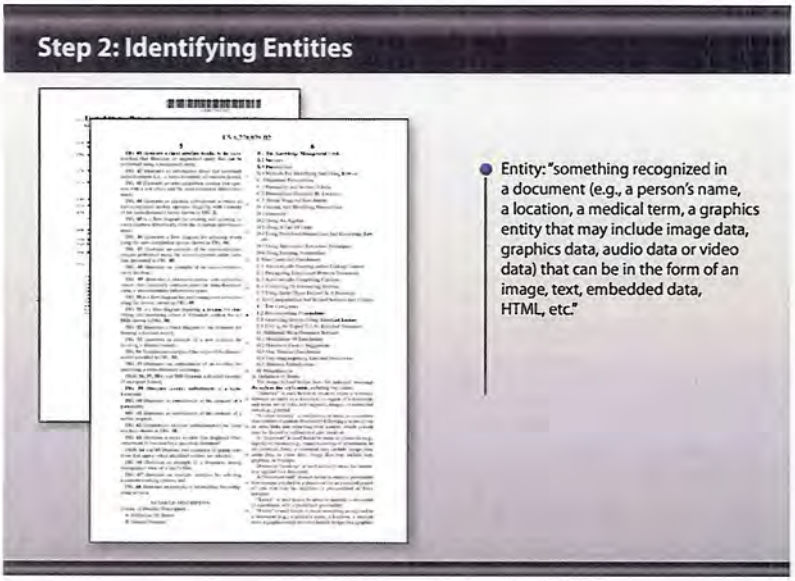


The example illustrated in figure 39 of the patent is similar to DMOZ – a nested hierarchy of categories (or "nodes") that becomes increasingly specific lower in the tree structure: Genetics is a subcategory of biology, which is itself a subcategory of science, which itself is a subcategory of the "toplevel" or "root" node.

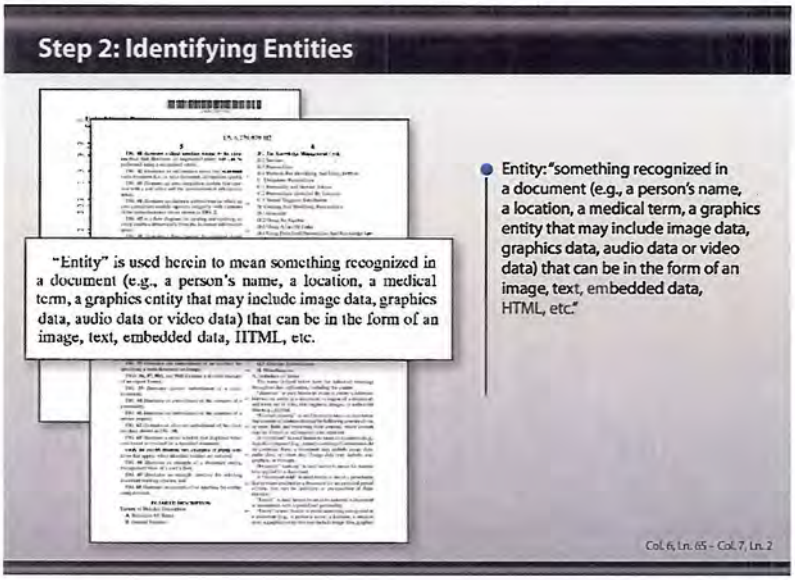


Note how the category is used to formulate a query to restrict a search. Here, the search for "seven up" has been restricted to the category "Top/Science/Biology/Genetics." This ensures that a search concerning "seven up" is done for documents in the category "genetics" rather than in a category for soft drinks.

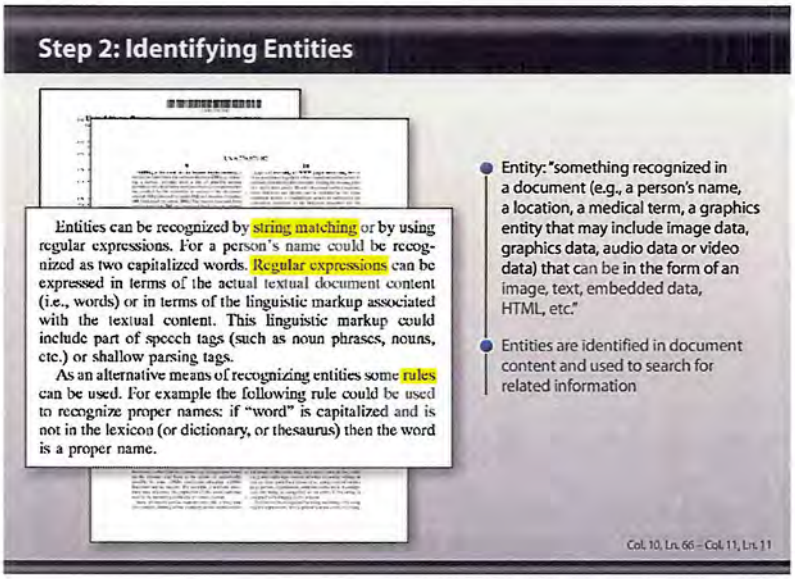
Step 2: Identifying Entities



After the claimed system has its defined organized classification, it can formulate queries concerning the subject matter of document content to restrict a search. The system scans document content for "entities," which are meant to represent concepts that might interest the user.



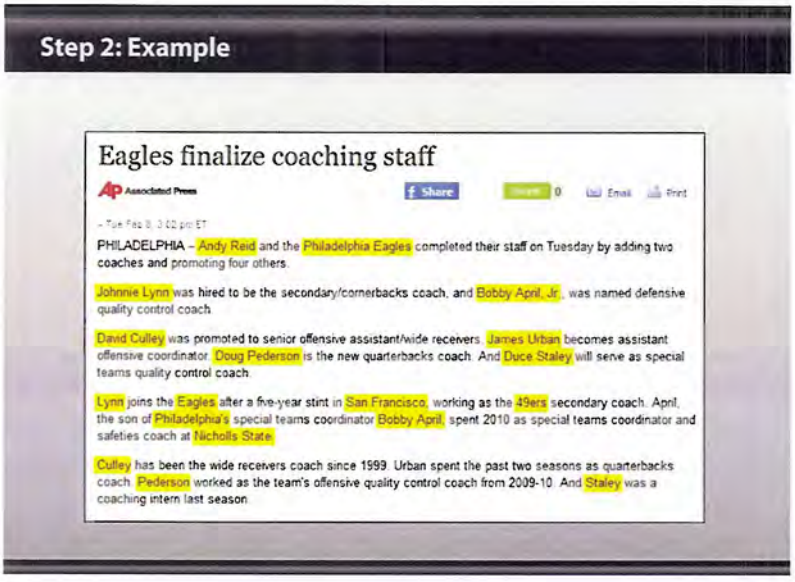
"Entity" is defined in the specification as "something recognized in a document" and may include names, places, or other proper nouns. For example, an entity may be the name of a person like "Michael Jordon" or "Ben Franklin," the name of a city like "Paris" or "Philadelphia," or the name of a product like "Toyota Prius" or "Furby."



The patent discusses various ways of identifying entities in a document, including string matching, regular expressions, and rules, all of which were known in the art.

Step 2: Example

For example, suppose a user were viewing a Web page containing a news article about the Philadelphia Eagles. The system might scan the article looking for entities, that is, words or phrases, that represent topics toward which it can direct searches.



In this case, the system might identify "Andy Reid," "Philadelphia Eagles," "Johnnie Lynn," and "Bobby April, Jr." as entities, among others.

Step 3: Categorizing to Assign a Classification Label to Document Content

Step 3: Categorizing to Assign a Classification Label to Document Content

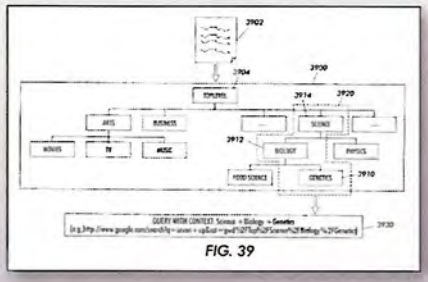


FIG. 39

- The system also assigns document content a classification label corresponding to a category from the organized classification

After identifying entities within the document content, the system assigns document content a classification label that corresponds to a category from the organized classification discussed earlier.

Step 3: Categorizing to Assign a Classification Label to Document Content

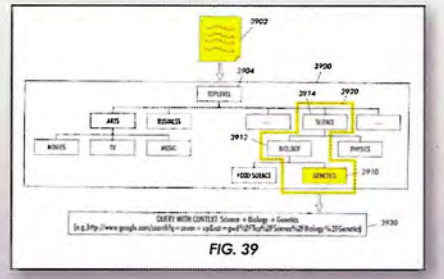


FIG. 39

- The system also assigns document content a classification label corresponding to a category from the organized classification

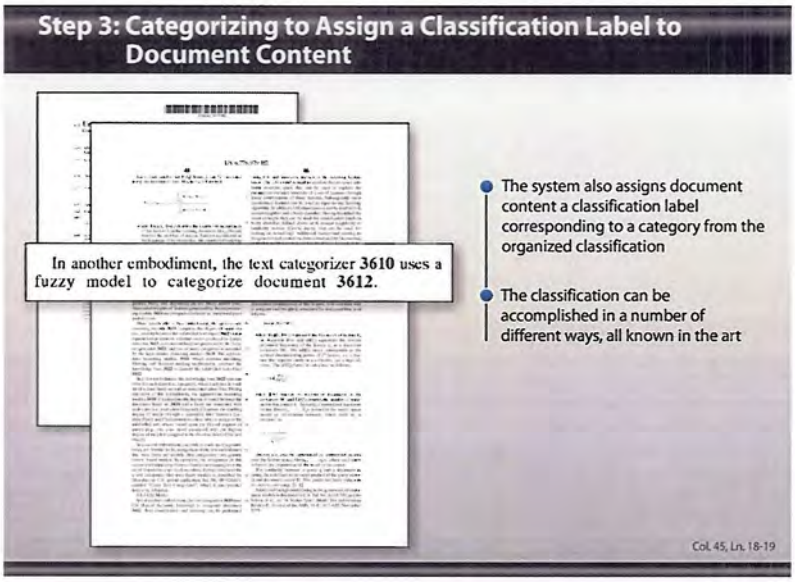
In the example given in figure 39, the document is given the classification label belonging to the subcategory "genetics."

Step 3: Categorizing to Assign a Classification Label to Document Content

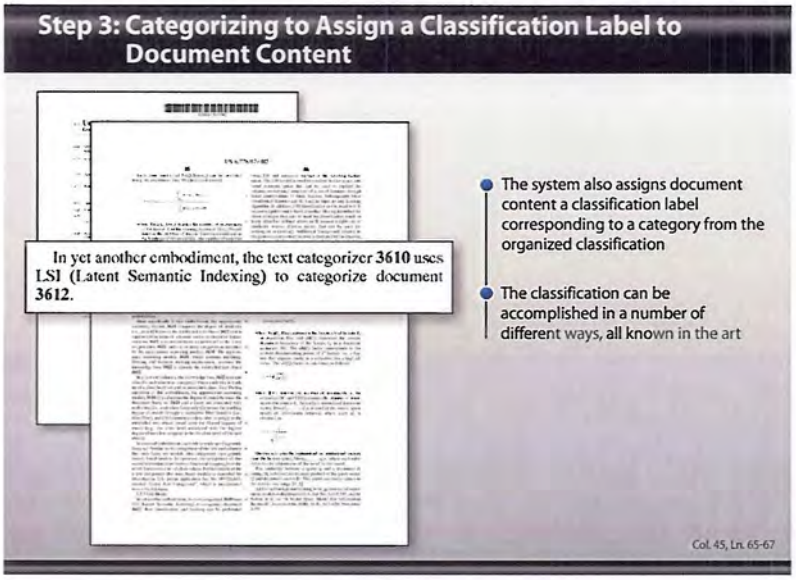
- The system also assigns document content a classification label corresponding to a category from the organized classification
- The classification can be accomplished in a number of different ways, all known in the art

The classification can be accomplished by various means known in the art at the time of the patent.

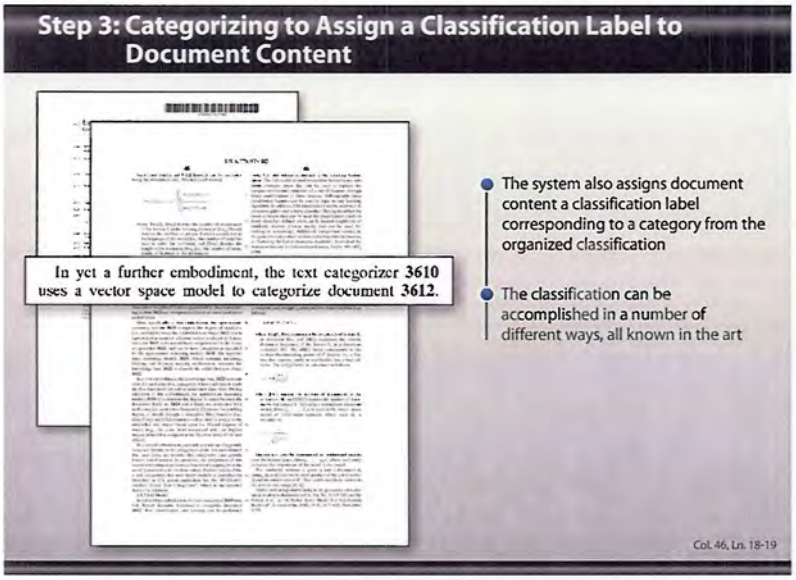
Step 3: Categorizing to Assign a Classification Label to Document Content



The patent discusses various such methods, for example, fuzzy models,

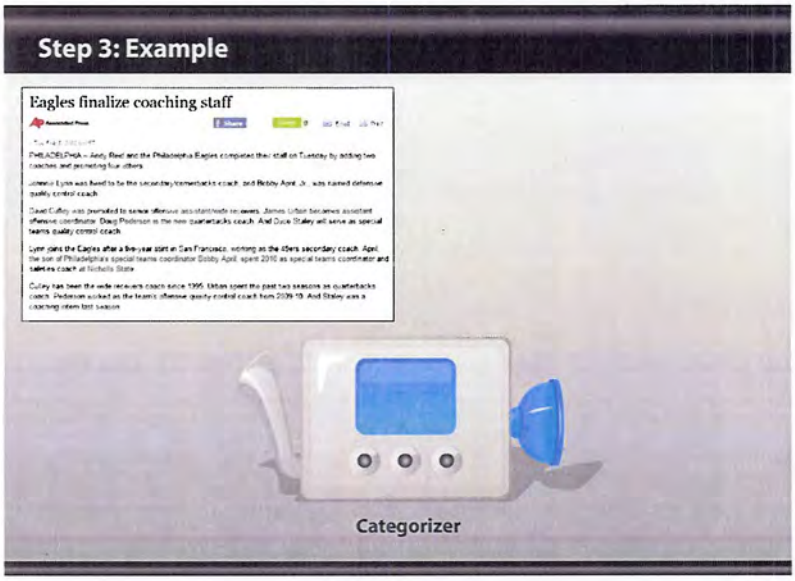


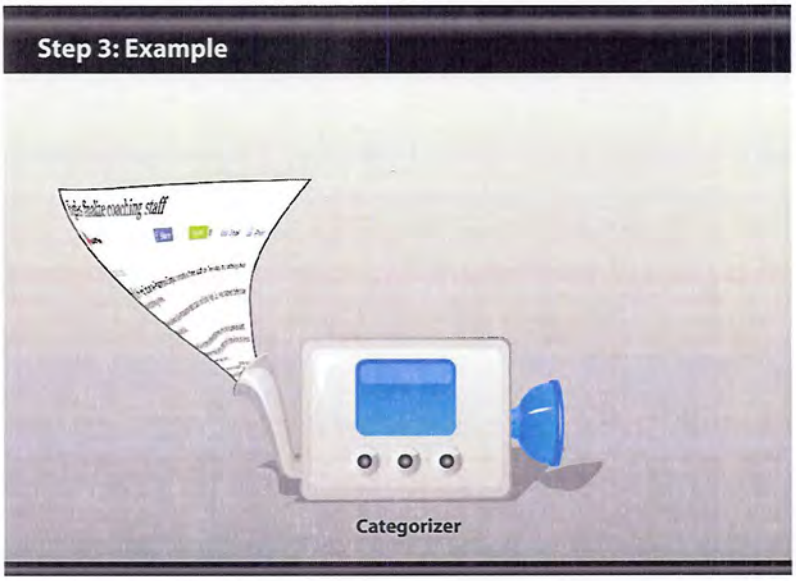
latent semantic indexing,



and a vector space model.

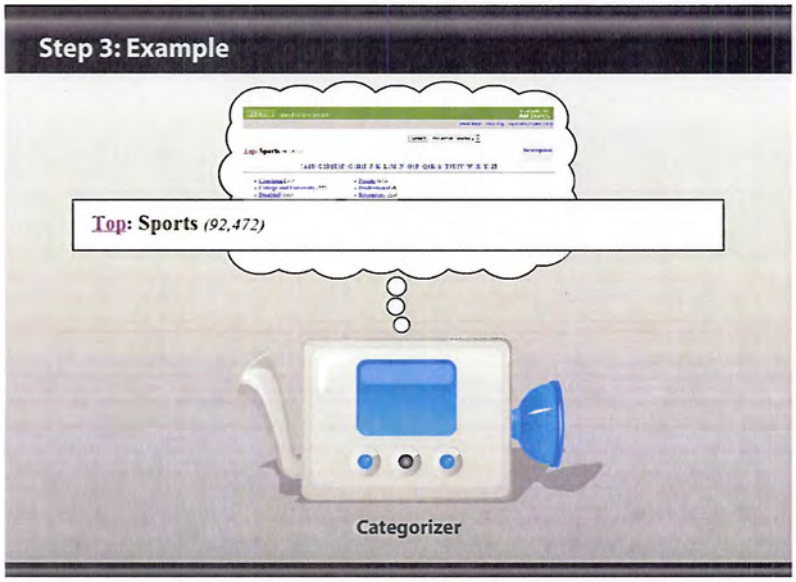
Step 3: Example



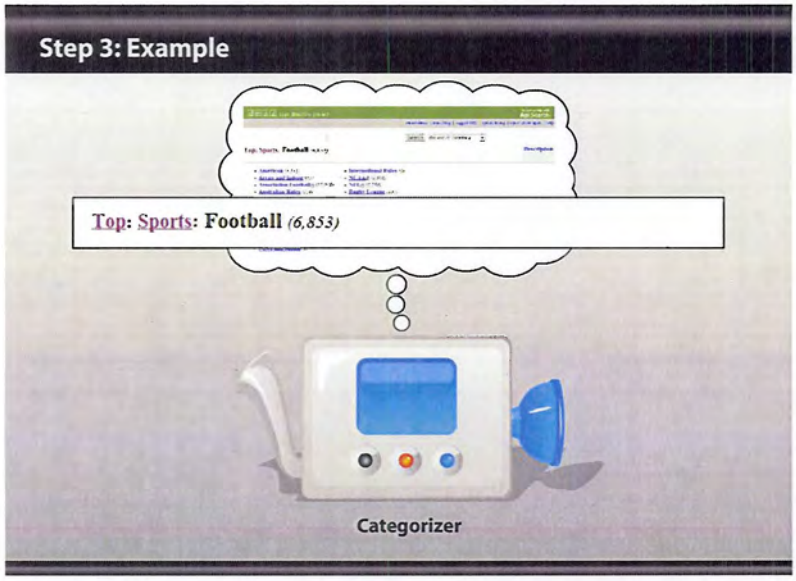


Going back to the example, the system would analyze content, then try to match the Eagles news article with one of the categories in the DMOZ hierarchy.

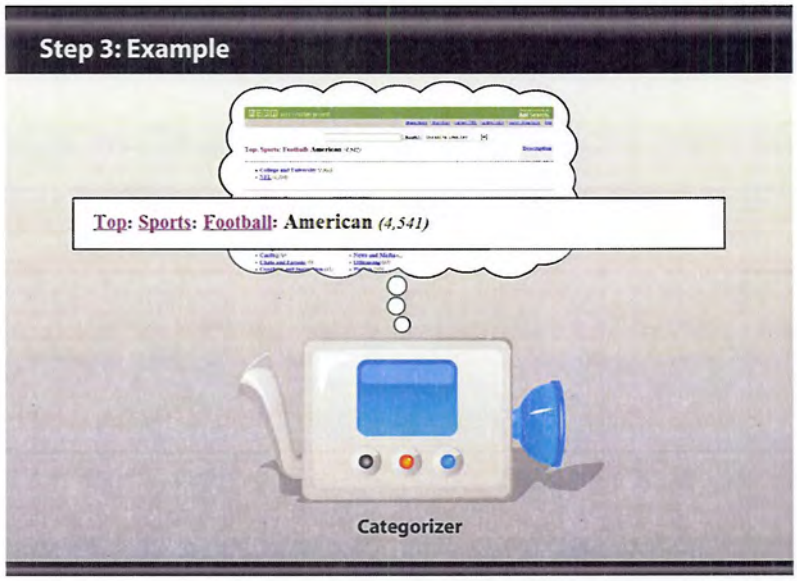
Step 3: Example



In this case, the categorizer might determine that the news article falls into the "Top," "Sports,"



"Football,"



"American,"