US 20040236739A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0236739 A1**

Nevill-Manning (43) **Pub. Date: Nov. 25, 2004**

(54) **SYSTEM AND METHOD FOR PROVIDING DEFINITIONS**

(76) Inventor: **Craig Nevill-Manning**, New York, NY (US)

Correspondence Address:
PATRICK J S INOUYE P S
810 3RD AVENUE
SUITE 258
SEATTLE, WA 98104 (US)

(21) Appl. No.: **10/608,270**

(22) Filed: **Jun. 27, 2003**

Related U.S. Application Data

(60) Provisional application No. 60/472,445, filed on May 20, 2003.

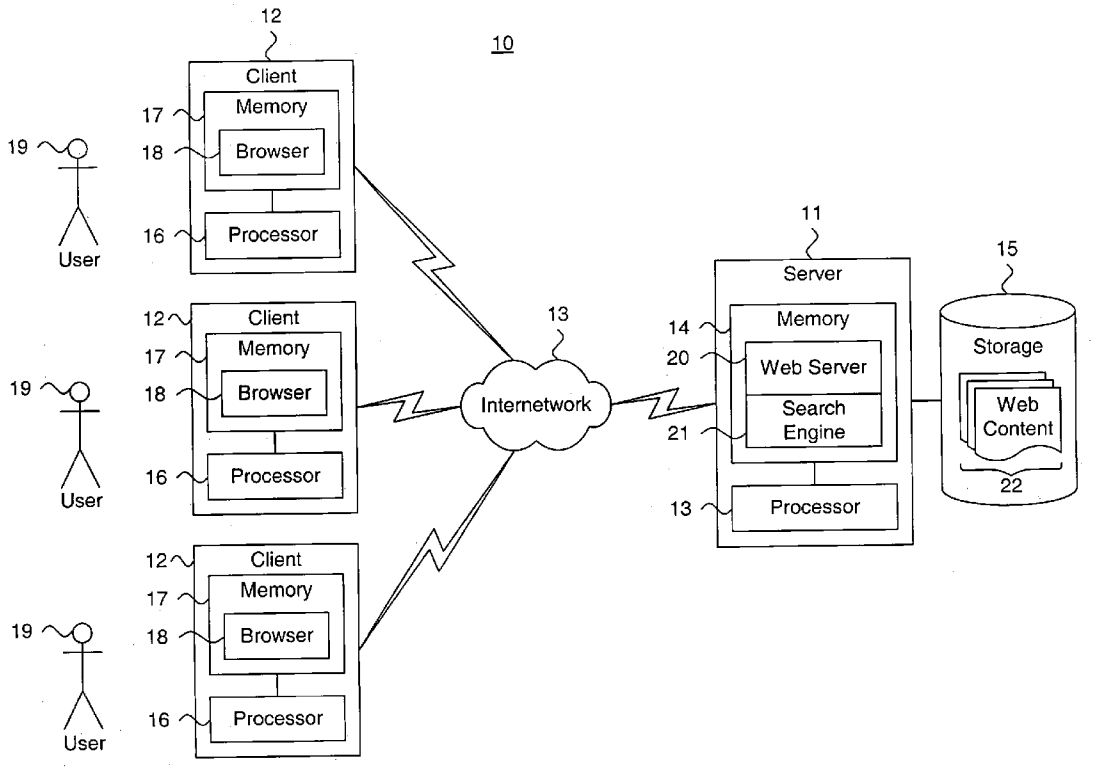
Publication Classification

(51) **Int. Cl.**[7] **G06F 7/00**

(52) **U.S. Cl.** **707/5**

(57) **ABSTRACT**

A system and method for providing definitions is described. A phrase to be defined is received. One or more documents, which each contain at least one definition, are determined. The phrase is matched to at least one of the definitions. One or more definitions for the phrase are presented.



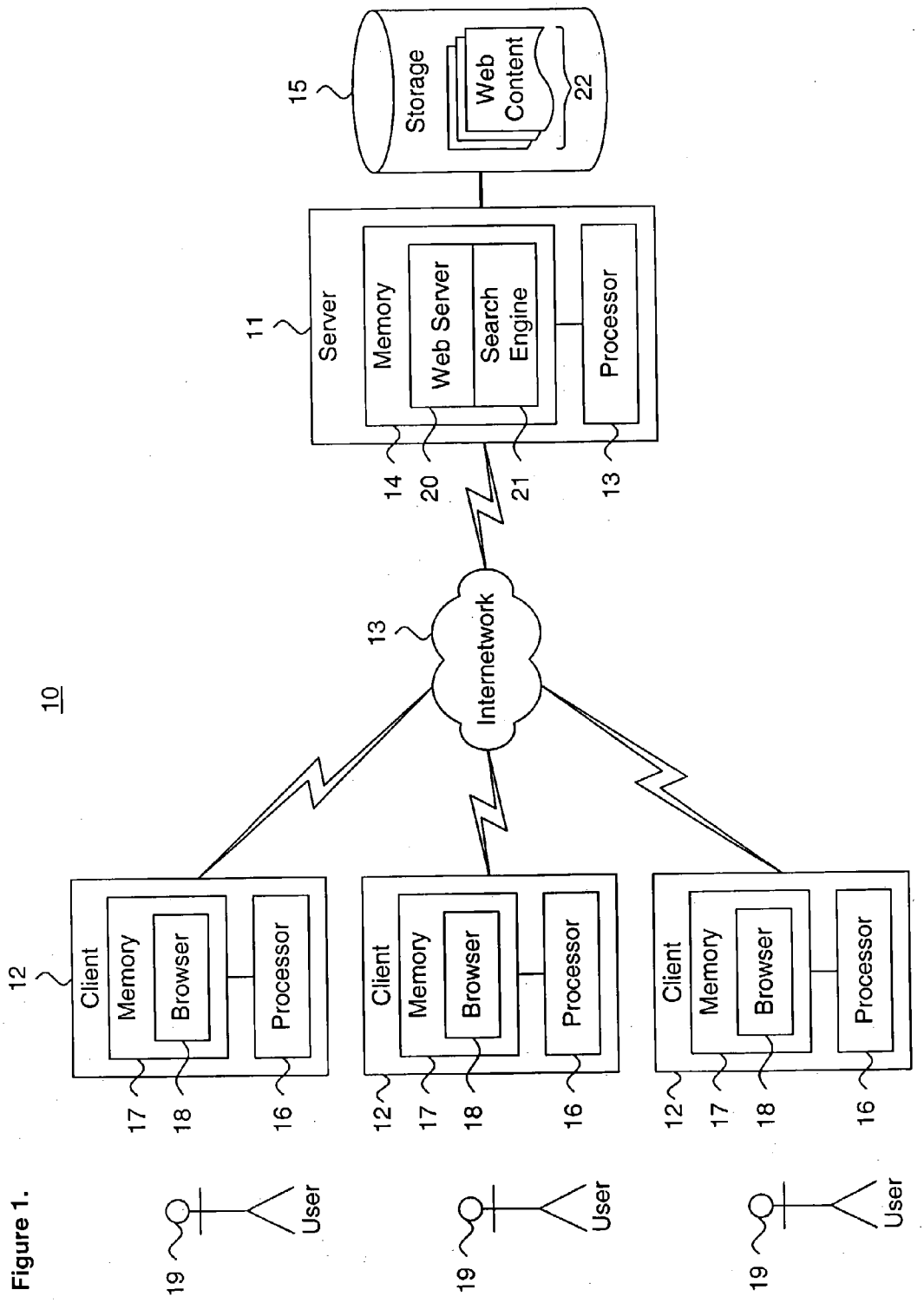

10
15
Storage
Web Content
22
11
Server
Memory
Web Server
Search Engine
Processor
14
20
21
13
13
Internetwork
12
Client
Memory
Browser
Processor
17
18
16
19
User

Figure 1.

Figure 2.

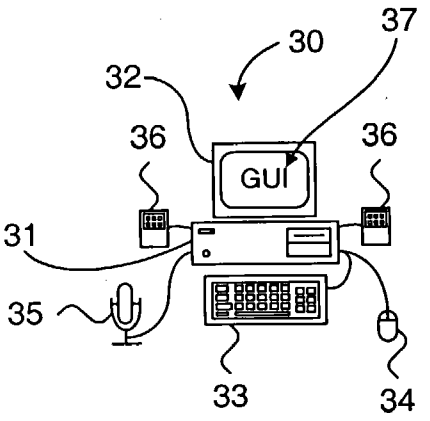


Figure 3.

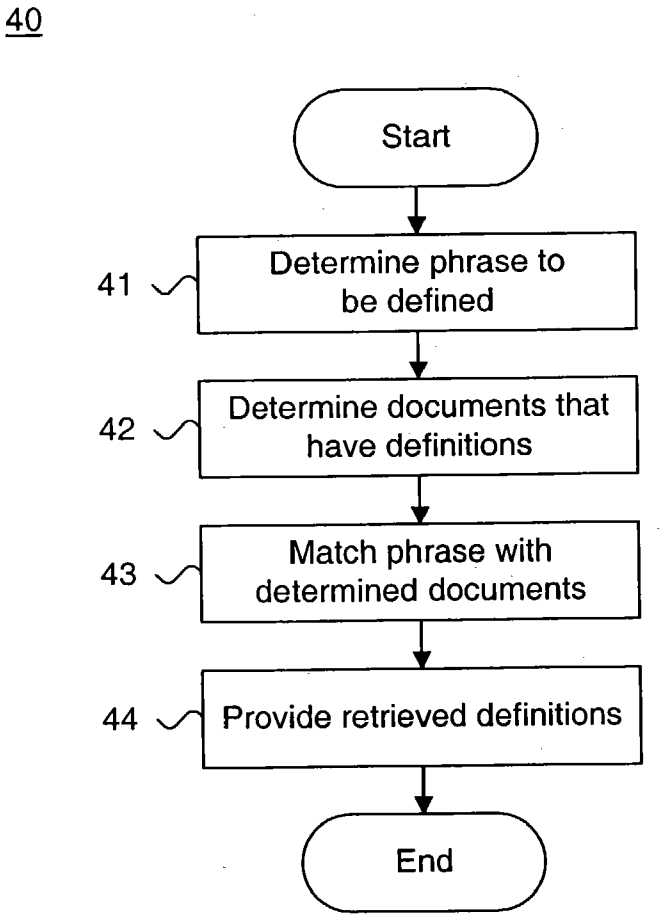


Figure 4.

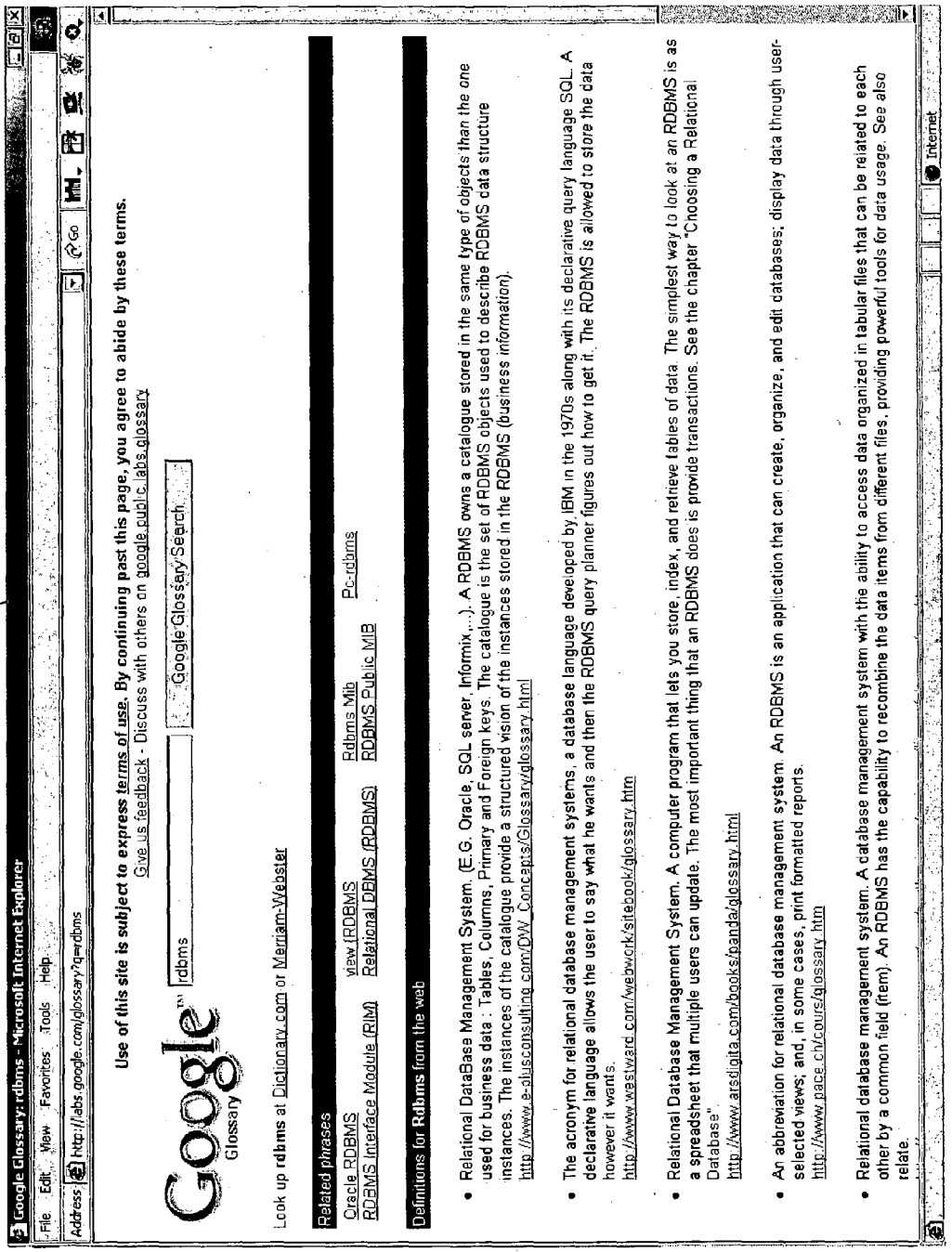

400
Google Glossary: rdbms - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://labs.google.com/glossary?q=rdbms
Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
Give us feedback - Discuss with others on google.public.labs.glossary
Google Glossary
rdbms
Google Glossary Search
Look up rdbms at Dictionary.com or Merriam-Webster
Related phrases
Oracle RDBMS
RDBMS Interface Module (RIM)
view (RDBMS
Relational DBMS (RDBMS)
Rdbms Mib
RDBMS Public MIB
Pc-rdbms
Definitions for Rdbms from the web
Relational DataBase Management System. (E.G. Oracle, SQL server, Informix,...). A RDBMS owns a catalogue stored in the same type of objects than the one used for business data : Tables, Columns, Primary and Foreign keys. The catalogue is the set of RDBMS objects used to describe RDBMS data structure instances. The instances of the catalogue provide a structured vision of the instances stored in the RDBMS (business information).
http://www.e-plusconsulting.com/DW_Concepts/Glossary/glossary.html
The acronym for relational database management systems, a database language developed by IBM in the 1970s along with its declarative query language SQL. A declarative language allows the user to say what he wants and then the RDBMS query planner figures out how to get it. The RDBMS is allowed to store the data however it wants.
http://www.westward.com/webwork/sitebook/glossary.htm
Relational Database Management System. A computer program that lets you store, index, and retrieve tables of data. The simplest way to look at an RDBMS is as a spreadsheet that multiple users can update. The most important thing that an RDBMS does is provide transactions. See the chapter "Choosing a Relational Database"
http://www.arstdigita.com/books/panda/glossary.html
An abbreviation for relational database management system. An RDBMS is an application that can create, organize, and edit databases; display data through user-selected views; and, in some cases, print formatted reports.
http://www.pace.ch/cours/glossary.htm
Relational database management system. A database management system with the ability to access data organized in tabular files that can be related to each other by a common field (item). An RDBMS has the capability to recombine the data items from different files, providing powerful tools for data usage. See also relate.
Internet

Figure 5.

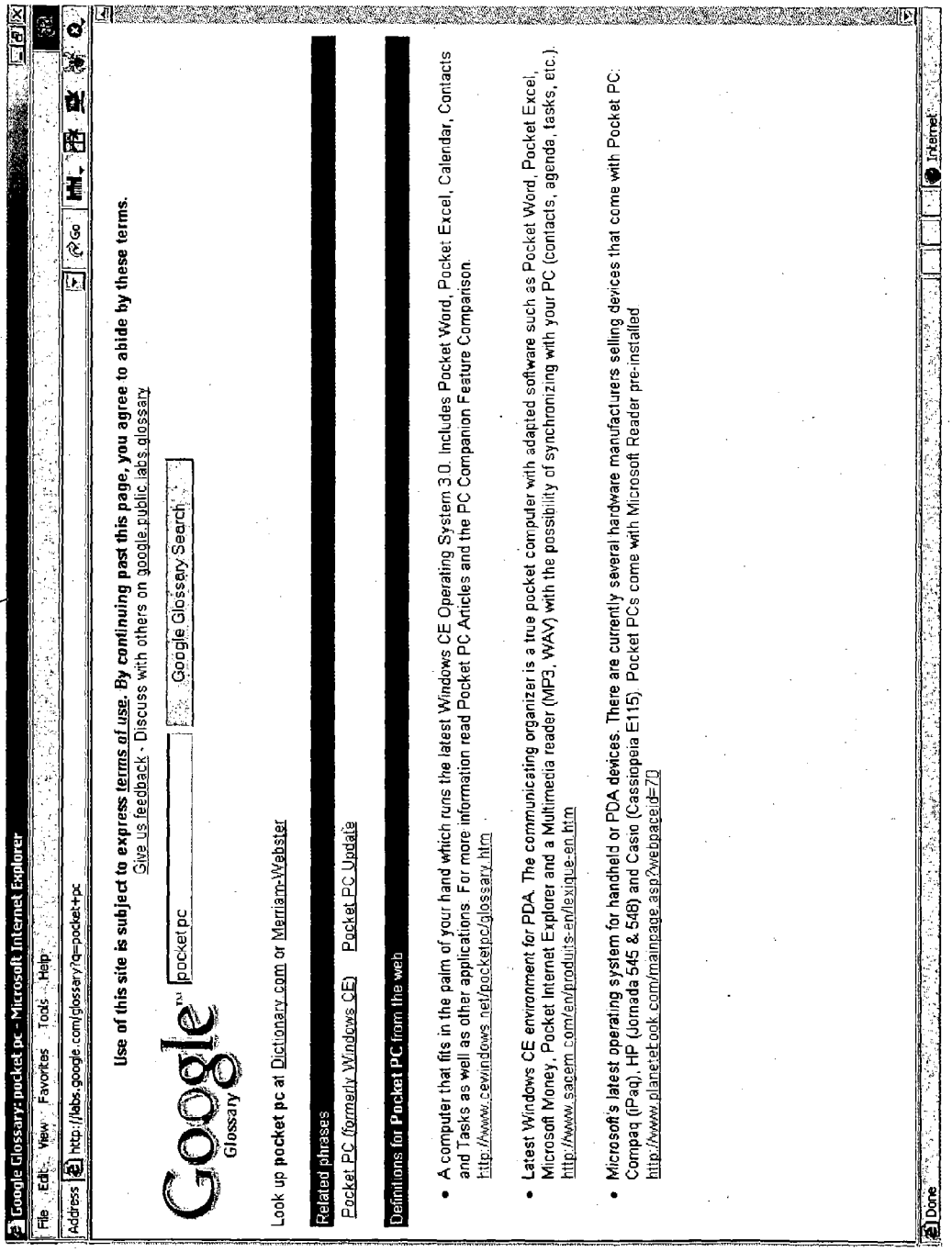
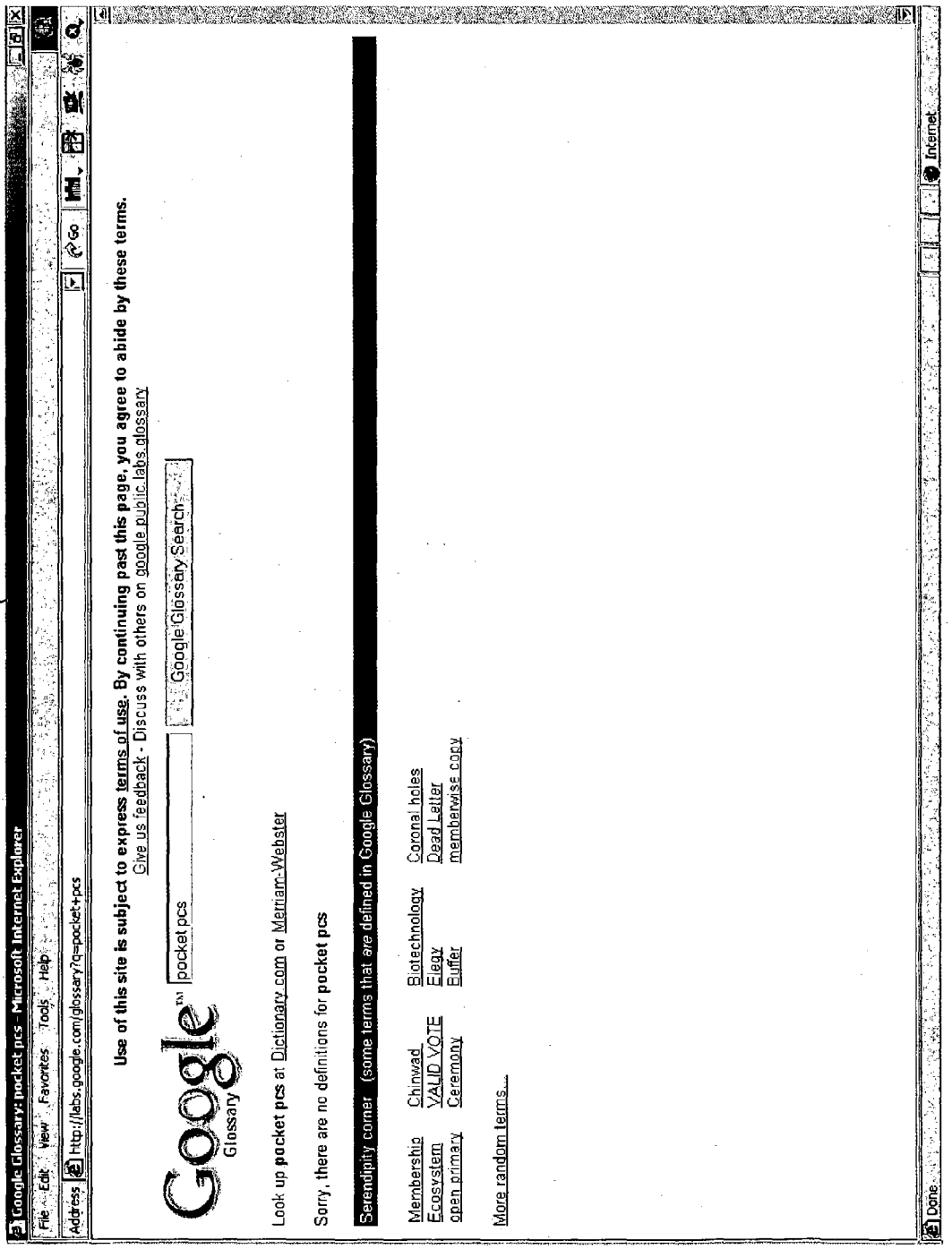

500
Google Glossary: pocket pc - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://labs.google.com/glossary?q=pocket+pc
Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
Give us feedback - Discuss with others on google.public.labs.glossary
Google Glossary
pocket pc
Google Glossary Search
Look up pocket pc at Dictionary.com or Merriam-Webster
Related phrases
Pocket PC (formerly Windows CE) Pocket PC Update
Definitions for Pocket PC from the web
A computer that fits in the palm of your hand which runs the latest Windows CE Operating System 3.0. Includes Pocket Word, Pocket Excel, Calendar, Contacts and Tasks as well as other applications. For more information read Pocket PC Articles and the PC Companion Feature Comparison.
http://www.cewindows.net/pocketpc/glossary.htm
Latest Windows CE environment for PDA. The communicating organizer is a true pocket computer with adapted software such as Pocket Word, Pocket Excel, Microsoft Money, Pocket Internet Explorer and a Multimedia reader (MP3, WAV) with the possibility of synchronizing with your PC (contacts, agenda, tasks, etc.).
http://www.sagem.com/en/produits-en/lexique-en.htm
Microsoft's latest operating system for handheld or PDA devices. There are currently several hardware manufacturers selling devices that come with Pocket PC: Compaq (iPaq), HP (Jornada 545 & 548) and Casio (Cassiopeia E115). Pocket PCs come with Microsoft Reader pre-installed
http://www.planetebook.com/mainpage.asp?webpageid=70
Done
Internet

Figure 6.


600
Google Glossary: pocket pcs - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://labs.google.com/glossary?q=pocket+pcs
Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
Give us feedback - Discuss with others on google public labs glossary
Google Glossary
pocket pcs
Google Glossary Search
Look up pocket pcs at Dictionary.com or Merriam-Webster
Sorry, there are no definitions for pocket pcs
Serendipity corner (some terms that are defined in Google Glossary)
Membership
Ecosystem
open primary
Chinwad
VALID VOTE
Ceremony
Biotechnology
Elegy
Buffer
Coronal holes
Dead Letter
memberwise copy
More random terms...
Done
Internet

SYSTEM AND METHOD FOR PROVIDING DEFINITIONS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This non-provisional patent application claims priority under 35 USC § 119(e) to U.S. provisional patent application, Ser. No. 60/472,445, filed May 20, 2003, the disclosure of which is incorporated by reference.

FIELD OF THE INVENTION

[0002] The present invention relates in general to providing definitions and, in particular, to a system and method for providing definitions.

BACKGROUND OF THE INVENTION

[0003] A system and method for providing definitions is described. There is a vast amount of content available on the Internet. Some of this content is organized in the form of glossaries or definitions. The system and methods described herein allow one to tap into these available resources to quickly and efficiently provide definitions for phrases. "Phrases" may refer to words, phrases, or any other semantic unit that is capable of definition.

SUMMARY OF THE INVENTION

[0004] An embodiment provides a system and method for providing definitions. A phrase to be defined is received. One or more documents, which each contain at least one definition, are determined. The phrase is matched to at least one of the definitions. One or more definitions for the phrase are presented.

[0005] A further embodiment provides determining definitions from distributed information stores. One or more documents are identified. Each document is maintained in a distributed information store and contains a definition for an associated phrase. Information regarding each identified document is stored. A phrase for which a definition is sought is matched against the stored information for each identified document. Each identified document is fetched from the distributed information store and one or more matching definitions are returned. Each matching definitions is presented.

[0006] Still other embodiments of the present invention will become readily apparent to those skilled in the art from the following detailed description, wherein are described embodiments of the invention by way of illustrating the best mode contemplated for carrying out the invention. As will be realized, the invention is capable of other and different embodiments and its several details are capable of modifications in various obvious respects, all without departing from the spirit and the scope of the present invention. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with the color drawings will be provided by the Office upon request and payment of the necessary fee.

[0008] **FIG. 1** is a block diagram showing a system for providing definitions, in accordance with the present invention.

[0009] **FIG. 2** is a block diagram showing a computer system for use in the system of **FIG. 1**.

[0010] **FIG. 3** is a flow diagram showing a method for providing definitions, in accordance with the present invention.

[0011] **FIG. 4** is a screen shot showing, by way of example, definitions provided by the method of **FIG. 3**.

[0012] **FIG. 5** is a screen shot showing, by way of example, further definitions provided by the method of **FIG. 3**.

[0013] **FIG. 6** is a screen shot showing, by way of example, still further definitions provided by the method of **FIG. 3**.

DETAILED DESCRIPTION

System Overview

[0014] **FIG. 1** is a block diagram showing a system **10** for providing definitions, in accordance with the present invention. A plurality of individual clients **12** are communicatively interfaced to a server **11** via an internetwork **13**, such as the Internet, or other form of communications network, as would be recognized by one skilled in the art. The individual clients **12** are operated by users **19** who transact requests for Web content and other operations through their respective client **12**.

[0015] In general, each client **12** can be any form of computing platform connectable to a network, such as the internetwork **13**, and capable of interacting with application programs. Exemplary examples of individual clients include, without limitation, personal computers, digital assistances, "smart" cellular telephones and pagers, lightweight clients, workstations, "dumb" terminals interfaced to an application server, and various arrangements and configurations thereof, as would be recognized by one skilled in the art. The internetwork **13** includes various topologies, configurations, and arrangements of network interconnectivity components arranged to interoperatively couple with enterprise, wide area and local area networks and include, without limitation, conventionally wired, wireless, satellite, optical, and equivalent network technologies, as would be recognized by one skilled in the art.

[0016] For Web content exchange and, in particular, to transact searches, each client **12** executes a Web browser **18** ("Web browser"), which implements a graphical user interface and through which search queries are sent to a Web server **20** executing on the server **11**, as further described below with reference to **FIG. 2**. Each search query describes or identifies information, generally in the form of Web content, which is potentially retrievable via the Web server **20**. In addition, the search query can include a phrase for which a definition is sought, as further described below with reference to **FIG. 3**. The search query provides characteristics, typically expressed as terms, such as keywords and the like, and attributes, such as language, character encoding and so forth, which enables a search engine **21**, also executing on the server **11**, to identify and send back Web pages.

The terms and attributes are a form of metadata, which constitute data describing data. Other styles, forms or definitions of search queries, search query characteristics, and metadata are feasible, as would be recognized by one skilled in the art.

[0017] The Web pages are sent back to the Web browser **18** for presentation, usually in the form of Web content titles, hyperlinks, and other descriptive information, such as snippets of text taken from the Web pages. The user can view or access the Web pages on the graphical user interface and can input selections and responses in the form of typed text, clicks, or both. The server **11** maintains an attached storage device **15** in which Web content **22** is maintained. The Web content **22** could also be maintained remotely on other Web servers (not shown) interconnected either directly or indirectly via the intemetwork **13** and which are preferably accessible by each client **12**.

[0018] The search engine **21** preferably identifies the Web content **22** best matching the search query terms to provide high quality Web pages, such as described in S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Search Engine" (1998) and in U.S. Pat. No. 6,285,999, issued Sep. 4, 2001 to Page, the disclosures of which are incorporated by reference. In identifying matching Web content **22**, the search engine **21** operates on information characteristics describing potentially retrievable Web content, as further described below with reference to **FIG. 2**. Note the functionality provided by the server **20**, including the Web server **20** and search engine **21**, could be provided by a loosely- or tightly-coupled distributed or parallelized computing configuration, in addition to a uniprocessing environment.

[0019] The individual computer systems, including server **11** and clients **12**, include general purpose, programmed digital computing devices consisting of a central processing unit (processors **13** and **16**, respectively), random access memory (memories **14** and **17**, respectively), non-volatile secondary storage **15**, such as a hard drive or CD ROM drive, network or wireless interfaces, and peripheral devices, including user interfacing means, such as a keyboard and display. Program code, including software programs, and data is loaded into the RAM for execution and processing by the CPU and results are generated for display, output, transmittal, or storage. The Web browser **18** is an HTTP-compatible Web browser, such as the Internet Explorer, licensed by Microsoft Corporation, Redmond, Wash.; Navigator, licensed by Netscape Corporation, Mountain View, Calif.; or a Mozilla or JavaScript enabled browser, as are known in the art.

Computer System Components

[0020] **FIG. 2** is a block diagram showing a computer system **30** for use in the system **10** of **FIG. 1**. The computer system **30** includes a processor **31** and visual display **32**, such as a computer monitor or liquid crystal diode (LCD) display, as are known in the art. The computer system **30** executes a Web browser **18** (shown in **FIG. 1**), which implements a graphical user interface **37**. Visual Web content, including retrieved definitions, is output within a display area defined on the graphical user interface **37** while user inputs are generally input both within the display area and within specified user input regions. Textual user inputs are received via a keyboard **33**. Linear, non-textual inputs are received via a pointing device **34**, such as a mouse, trackball, track pad, or arrow keys. Similarly, voice- and sound-based inputs are received via a microphone **35**. Visual outputs are displayed via the graphical user interface **37** on the visual display **32**, while audio outputs are played on the speakers **36**. Other forms of computer components, including processor **31**, visual display **32**, and input and output devices could be used, as would be recognized by one skilled in the art.

Method Overview

[0021] One embodiment of the present invention will now be described with reference to **FIG. 3**, which provides a flow diagram showing a method for providing definitions, in accordance with the present invention. The method is described as a sequence of process operations or steps, which can be executed, for instance, by the system of **FIG. 1**, or equivalent component.

[0022] First, a phrase for which definition is sought is provided (block **310**). The phrase may be provided by, for example, a user request or query, or by any other means. One example of a system for providing a phrase is that located at the URL identified by http://labs.google.com/glossary, the contents of which are incorporated by reference. In addition, the spelling of the phrase can be corrected if necessary or normalized into a common root form to provide more consistent definition results.

[0023] Documents that contain definitions are determined (block **320**). These documents may be determined in any number of ways. For example, such documents may be determined during Web-crawling or spidering performed by search engines in either real time or batch processing modes. Once a document is determined to contain definitions, the document (or information about the document, such as the document's URL) may be stored or remembered for future use. "Authoritative" sources for definitions may also be used, for example, documents associated with Web sites, such as http://www.dictionary.com.

[0024] In one embodiment of the present invention, documents containing definitions are located substantially in real time, by conducting a query via an Internet search engine. In a further embodiment, the documents are located substantially in a batch processing mode, for example, by fetching, parsing and indexing the documents containing definitions off-line prior to receiving queries. In addition, a combination could be used, such as by providing batch processing for identifying documents containing definitions and using real time processing to fetch, de-duplicate and clean up definitions responsive to each query.

[0025] The query may search for terms that are likely to indicate the presence of definitions, such as "glossary," "definition," "dictionary," and so forth, as well as variants and canonicalizations thereof. The search may be conducted over the document text as a whole, or may be restricted to certain portions or fields within documents, such as the title field, fields containing other metadata, and so forth. The structure of documents, that is, the tagged nature of HTML documents, may also be relevant to determining how to structure the query. In an embodiment of the invention, a search for "glossary," "definitions," or "dictionary" in the title of Web pages are used to determine the relevant documents. As will be recognized by one of ordinary skill in

the art of information retrieval, the above methods may be combined in various fashions and with numerous other methods to determine definition containing documents.

[0026] The phrase for which definition is sought is then matched against the determined documents to return definitions (block **330**). The documents determined in this step (block **330**) may be parsed to identify occurrences of the phrase being sought and the phrase's associated definition. For example, definition containing documents may be organized with "headwords," or words that can be looked up in a dictionary form. There are various methods for identifying headwords and/or identifying definitions. In one embodiment of the invention, one or more of the following methods are used to parse apart documents, identify headwords, and/or return definitions:

[0027] If the page uses <dl>, <dt> and <dd>, which are HTML tags used for specifying lists of definitions, the HTML mark up is relied upon to identify definitions, that is:

```
An example definition list
<dl>
  <dt>Headword 1
    <dd>This is the definition of Headword 1
  <dt>Headword 2
    <dd>This is the definition of Headword 2
  <dt>Headword 3
    <dd>This is the definition of Headword 3
</dl>
```

[0028] HTML tags, such as <p>, <tr>, <li>, and
, may be treated as separators between successive definitions.

[0029] White space or punctuation (.,;-) is eliminated at the beginning of definitions.

[0030] Headwords may be identified by the fact that the headwords are surrounded by the HTML tags <b>, , , <code>, or .

[0031] Lines that do not start with headwords are deleted.

[0032] If there are fewer than N, for instance, N=5, definitions found in the document or page, all definitions in the document or page are discarded.

[0033] The parser does not need to be perfect at identifying all headwords and definitions. In one embodiment, due to the large number of definition-containing documents determined in the definition document determination step (block **320**), the parser is biased towards precision rather than thoroughness. In other words, the parser errs towards throwing entries away rather than keeping entries that may be incorrect because there are more than enough definitions to supply a satisfactory outcome. Similarly, in a further embodiment, the parser de-duplicates entries that are duplicative or merely cumulative of other entries.

[0034] One or more of the returned definitions are then provided (block **340**). In one embodiment, the returned definitions are ranked according to PageRank™ of the documents from which they are retrieved, according to the methods disclosed in U.S. Pat. No. 6,285,999, cited above. The retrieved definitions may also be processed for presentation, such as by carrying out one or more of the following steps:

[0035] Removing:

[0036] all HTML markup;

[0037] leading and trailing white space in both headword and definition;

[0038] all punctuation: (.:;!?-) in the headword;

[0039] all leading non-alpha and non-parenthesis in the headword and definition;

[0040] all trailing non-alphanumeric and non-parenthesis in the headword.

[0041] Throw the definition away if:

[0042] the definition starts with "see"

[0043] the definition is a duplicate of one already retrieved.

[0044] Capitalize the first letter the definition.

[0045] In one embodiment, only definitions whose head phrases are an exact match for the phrase are presented. However, in other embodiments of the invention, a looser form of matching may be allowed.

[0046] Other information may also be determined and presented. In one embodiment of the present invention, superstrings of the phrase are tabulated and presented as query refinements or related phrases. Superstrings are strings that contain the phrase (or possibly common variants or canonicalized versions of the phrase). Methods for determining common variants or canonicalized versions of words and phrases are described in, for example, U.S. patent application Ser. No. 10/377,117, Attorney Docket No. GP-091-00-US, entitled "SEARCH QUERIES IMPROVED BASED ON QUERY SEMANTIC INFORMATION," filed Mar. 3, 2003, pending, and listing Amit Singhal et al. as inventors, which disclosure is incorporated by reference. For example, the top M superstrings may be listed. Similarly, the phrase may be presented in a processed form, such as in the phrase's most common capitalization; for instance, a user query for [pocket pc] or [pocket pcs] may be presented as "Pocket PC" because that is the most common form and/or capitalization found in the definitions.

[0047] As will be recognized by one of skill in the art, the steps described above with reference to **FIG. 3** need not be performed in the order listed, and steps may be added or removed.

[0048] As used in this specification, a "document" is to be broadly interpreted to include any machine readable or machine storable work product. A document may be a file, a combination of files, one or more files with embedded links to other files, and so forth. The files may be of any type, such as text, audio, image, video, and so forth. In the context of the Internet, a common document is a Web page, as is known in the art.

[0049] According to a further aspect of the invention, in situations where no definitions are found (or where definitions are not selected for presentation, such as if there is doubt as to whether the definition properly matches the original provided phrase), a set of terms or phrases that are

related to the original phrase, that are deemed likely to be related to the phrase, that may be of interest (e.g. of interest to the user entering the original phrase), or even a "random" or eclectic set of terms or phrases for which definitions are returned, may be provided. Such terms may be provided, for example, to give a user a guide as to the types of terms that are defined, or for user amusement.

Sample Web Pages

[0050] **FIG. 4** is a screen shot **400** showing, by way of example, definitions provided by the method of **FIG. 3**. A glossary search for the phrase "rdbms" is provided, substantially as shown.

[0051] **FIG. 5** is a screen shot **500** showing, by way of example, further definitions provided by the method of **FIG. 3**. A glossary search for the phrase "pocket pc" is provided, substantially as shown.

[0052] **FIG. 6** is a screen shot **600** showing, by way of example, still further definitions provided by the method of **FIG. 3**. A glossary search for the phrase "pocket pcs" is provided, substantially as shown.

[0053] While the invention has been particularly shown and described as referenced to the embodiments thereof, those skilled in the art will understand that the foregoing and other changes in form and detail may be made therein without departing from the spirit and scope of the invention.

What is claimed is:

1. A system for providing definitions, comprising:

a server receiving a phrase to be defined, determining one or more documents each containing at least one definition, and matching the phrase to at least one of the definitions; and

a user interface presenting one or more definitions for the phrase.

2. The system of claim 1, wherein receiving the phrase to be defined, determining one or more documents each containing at least one definition, matching the phrase to at least one of the definitions, and presenting one or more definitions for the phrase are performed substantially in real time, batch mode, or a combination thereof.

3. The system of claim 1, wherein the documents are Web pages.

4. The system of claim 1, wherein the determining includes conducting a query on a search engine.

5. The system of claim 4, wherein the determining includes searching for documents that include a predetermined term in a predetermined field.

6. The system of claim 5, wherein the predetermined term includes one of a glossary, definition, and dictionary.

7. The system of claim 5, wherein the predetermined field is a title field.

8. The system of claim 1, wherein the matching includes determining the presence of the phrase in one or more determined documents.

9. The system of claim 8, wherein the matching includes determining the absence of the phrase in one or more determined documents.

10. The system of claim 8, wherein determining the presence of the phrase further includes determining an exact match of the phrase.

11. The system of claim 8, wherein the matching comprises modifying the phrase.

12. The system of claim 8, wherein modifying the phrase comprises determining a canonical form of the phrase.

13. The system of claim 1, wherein the matching further comprises retrieving an associated definition of the phrase.

14. The system of claim 1, wherein presenting one or more definitions includes ranking the definitions.

15. The system of claim 14, wherein the ranking is based at least in part on the documents.

16. The system of claim 15, wherein the ranking is based at least in part on the PageRank of the documents associated with the definitions.

17. The system of claim 1, wherein the presenting further includes processing the definitions.

18. The system of claim 1, wherein presenting definitions for the phrase includes presenting a substantially most common capitalization of the phrase.

19. The system of claim 18, further comprising presenting less common forms of the phrase.

20. The system of claim 1, further comprising determining superstrings of the phrase present in the documents.

21. The system of claim 20, further comprising presenting at least some of the determined superstrings.

22. The system of claim 21, wherein at least one of the presented superstrings is presented as one of a related phrase and a suggested query.

23. A method for providing definitions, comprising:

receiving a phrase to be defined;

determining one or more documents each containing at least one definition;

matching the phrase to at least one of the definitions; and

presenting one or more definitions for the phrase.

24. The method of claim 23, wherein receiving the phrase to be defined, determining one or more documents each containing at least one definition, matching the phrase to at least one of the definitions, and presenting one or more definitions for the phrase are performed substantially in real time, batch mode, or a combination thereof.

25. The method of claim 23, wherein the documents are Web pages.

26. The method of claim 23, wherein the determining includes conducting a query on a search engine.

27. The method of claim 23, wherein the determining includes searching for documents that include a predetermined term in a predetermined field.

28. The method of claim 27, wherein the predetermined term includes one of a glossary, definition, and dictionary.

29. The method of claim 27, wherein the predetermined field is a title field.

30. The method of claim 23, wherein the matching includes determining the presence of the phrase in one or more determined documents.

31. The method of claim 30, wherein the matching includes determining the absence of the phrase in one or more determined documents.

32. The method of claim 30, wherein determining the presence of the phrase further includes determining an exact match of the phrase.

33. The method of claim 30, wherein the matching comprises modifying the phrase.

34. The method of claim 30, wherein modifying the phrase comprises determining a canonical form of the phrase.

35. The method of claim 23, wherein the matching further comprises retrieving an associated definition of the phrase.

36. The method of claim 23, wherein presenting one or more definitions includes ranking the definitions.

37. The method of claim 36, wherein the ranking is based at least in part on the documents.

38. The method of claim 37, wherein the ranking is based at least in part on the PageRank of the documents associated with the definitions.

39. The method of claim 23, wherein the presenting further includes processing the definitions.

40. The method of claim 23, wherein presenting definitions for the phrase includes presenting a substantially most common capitalization of the phrase.

41. The method of claim 40, further comprising presenting less common forms of the phrase.

42. The method of claim 23, further comprising determining superstrings of the phrase present in the documents.

43. The method of claim 42, further comprising presenting at least some of the determined superstrings.

44. The method of claim 43, wherein at least one of the presented superstrings is presented as one of a related phrase and a suggested query.

45. A computer-readable storage medium holding code for performing the method according to claim 23.

46. An apparatus for providing definitions, comprising:

means for receiving a phrase to be defined;

means for determining one or more documents each containing at least one definition;

means for matching the phrase to at least one of the definitions; and

means for presenting one or more definitions for the phrase.

47. A system for determining definitions from distributed information stores, comprising:

a search engine identifying one or more documents, which is each maintained in a distributed information store and contains a definition for an associated phrase, and storing information regarding each identified document; and

a search front end matching a phrase for which a definition is sought against the stored information for each identified document, fetching each identified document from the distributed information store and returning one or more matching definitions, and presenting each matching definitions.

48. A system according to claim 47, further comprising:

a repository storing the information for a subset of the identified documents.

49. A system according to claim 47, further comprising:

a query engine conducting a query for the phrase for which a definition is sought, comprising at least one of searching for at least one of terms, phrases, variants, and canonicalizations indicating a presence of a definition, searching for text or fields within a document indicating a presence of a definition, and searching a structure of a document indicating a presence of a definition.

50. A system according to claim 47, further comprising:

a parser parsing the identified documents to identify occurrences of the phrase for which a definition is sought.

51. A system according to claim 47, further comprising:

a processor processing the matching definitions, comprising at least one of:

a filter limiting the matching definitions to substantially matching definitions; and

a definitions module providing at least one of a superstring, common variants, and common forms of the phrase for which a definition is sought.

52. A system according to claim 47, wherein the matching definitions comprise at least one of matching terms and phrases, related terms and phrases, and random and eclectic terms and phrases.

53. A method for determining definitions from distributed information stores, comprising:

identifying one or more documents, which is each maintained in a distributed information store and contains a definition for an associated phrase, and storing information regarding each identified document;

matching a phrase for which a definition is sought against the stored information for each identified document;

fetching each identified document from the distributed information store and returning one or more matching definitions; and

presenting each matching definitions.

54. A method according to claim 53, further comprising:

storing the information for a subset of the identified documents.

55. A method according to claim 53, further comprising:

conducting a query for the phrase for which a definition is sought, comprising at least one of:

searching for at least one of terms, phrases, variants, and canonicalizations indicating a presence of a definition;

searching for text or fields within a document indicating a presence of a definition; and

searching a structure of a document indicating a presence of a definition.

56. A method according to claim 53, further comprising:

parsing the identified documents to identify occurrences of the phrase for which a definition is sought.

57. A method according to claim 53, further comprising:

processing the matching definitions, comprising at least one of:

limiting the matching definitions to substantially matching definitions; and

providing at least one of a superstring, common variants, and common forms of the phrase for which a definition is sought.

58. A method according to claim 53, wherein the matching definitions comprise at least one of matching terms and phrases, related terms and phrases, and random and eclectic terms and phrases.

59. A computer-readable storage medium holding code for performing the method according to claim 53.

60. An apparatus for determining definitions from distributed information stores, comprising:

means for identifying one or more documents, which is each maintained in a distributed information store and contains a definition for an associated phrase, and means for storing information regarding each identified document;

means for matching a phrase for which a definition is sought against the stored information for each identified document;

means for fetching each identified document from the distributed information store and means for returning one or more matching definitions; and

means for presenting each matching definitions.

* * * * *