(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 October 2004 (21.10.2004)

PCT

(10) International Publication Number
WO 2004/090755 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number:
PCT/US2004/009766

(22) International Filing Date: 31 March 2004 (31.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/459,339 31 March 2003 (31.03.2003) US
10/407,476 3 April 2003 (03.04.2003) US

(71) Applicant *(for all designated States except US)*: **GOOGLE INC.** [US/US]; 2400 Bayshore Parkway, Mountain View, CA 94043 (US).

(72) Inventors: **LAMPING, John**; 1299 Eva Avenue, Los Altos, CA 94024 (US). **GOMES, Ben**; 79 Eldora, Mountain View, CA 94041 (US). **MCGRATH, Mizuki**; 175 Bluxome Street #316, San Francisco, CA 94107 (US). **SINGHAL, Amit**; 2435 Aztec Way, Palo Alto, CA 94303 (US).

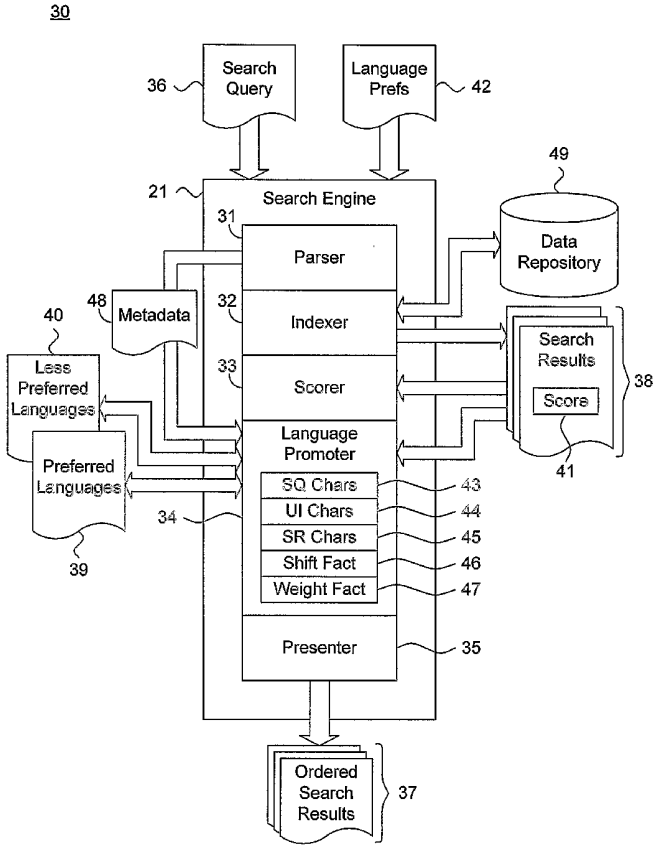
(74) Agent: **INOUYE, Patrick, Joseph, Sus**; Law Offices of Patrick J.S. Inouye, 810 Third Avenue, Suite 258, Seattle, WA 98104 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK,

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR PROVIDING PREFERRED LANGUAGE ORDERING OF SEARCH RESULTS



(57) Abstract: A system and method for providing preferred language (39) ordering (85) of search results (38) is described. A search query (36) describing potentially retrievable information (22) provided in a plurality of search result languages (37) is received. A search (36) is executed by evaluating the search query (36) against information characteristics maintained in a searchable data repository. At least one preferred language (39) applicable to search results (38) generated responsive to the executed search (36) is dynamically determined. At least some of the search results (38) are ordered in consideration of the at least one preferred language (39).

WO 2004/090755 A2

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEM AND METHOD FOR PROVIDING PREFERRED LANGUAGE ORDERING OF SEARCH RESULTS

CROSS-REFERENCE TO RELATED APPLICATION

This patent application claims priority under 35 USC § 119(e) to U.S. provisional patent application, Serial No.60/459,339, entitled "System And Method For Providing Preferred Language Ordering Of Search Results," filed March 31, 2003 and claims priority under 35 USC § 120 to the U.S. non-provisional patent application, Serial No. 10/407,476, filed April 3, 2003, the disclosures of which are incorporated by reference.

TECHNICAL FIELD

The present invention relates in general to information location and, in particular, to a system and method for providing preferred language ordering of search results.

BACKGROUND ART

Although the Internet traces back to the late 1960s, the widespread availability and acceptance of personal computing and internetworking have resulted in the explosive growth and unprecedented advances in information sharing technologies. In particular, the Worldwide Web ("Web") has revolutionized accessibility to untold volumes of information in stored electronic form to a worldwide audience, including written, spoken (audio) and visual (imagery and video) information, both in archived and real-time formats. In short, the Web has provided desktop access to every connected user to a virtually unlimited library of information in almost every language worldwide.

Information exchange on the Web operates under a client-server model. Individual clients execute Web content retrieval and presentation applications, typically in the form of Web browsers. The Web browsers send request messages for Web content to centralized Web servers, which function as data storage and retrieval repositories. The Web servers parse the request messages and return the requested Web content in response messages.

Search engines have evolved in tempo with the increased usage of the Web to enable users to find and retrieve relevant Web content in an efficient and timely manner. As the amount and types of Web content has increased, the sophistication and accuracy of search engines has likewise improved. Generally, search engines strive to provide the highest quality results in response to a search query. However, determining quality is difficult, as the relevance of

retrieved Web content is inherently subjective and dependent upon the interests, knowledge and attitudes of the user.

Existing methods used by search engines are based on matching search query terms to terms indexed from Web pages. More advanced methods determine the importance of retrieved Web content using, for example, a hyperlink structure-based analysis, such as described in S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Search Engine," (1998) and in U.S. Patent No. 6,285,999, issued September 4, 2001 to Page, the disclosures of which are incorporated by reference.

A typical search query scenario begins with either a natural language question or individual keywords submitted to a search engine. The search engine executes a search against a data repository describing information characteristics of potentially retrievable Web content and identifies the candidate search results. Searches can often return thousands or even millions of results, so most search engines typically rank or score only a subset of the most promising results. Targeted search results can also be introduced, such as advertising or topical information content. The top search results are then presented to the user, usually in the form of Web content titles, hyperlinks, and other descriptive information, such as snippets of text taken from the search results.

Search engines are generally available to users located worldwide. Thus, part of providing high-quality search results is being able to provide those search results in languages acceptable to the requesting user. Acceptable languages include languages specified by the user, as well as other acceptable languages. For instance, a French-preferring user might also accept search results in English. Acceptable languages can also include related languages and dialects. For example, Portuguese search results might be acceptable to a user who generally prefers Spanish. Finally, acceptable languages can include dead languages, such as classical Greek or Olde English, or psuedo-languages, such as Klingon. Dead and psuedo-languages are typically not supported by search engines, but may nevertheless reflect the academic, historic, or personal interests of the requesting user.

Currently, the Hypertext Transfer Protocol (HTTP) is used by most Web browser, Web server, and related Web applications, to transact Web information exchange. HTTP is a session-less protocol and no state identifying user preferences, including language, is typically maintained. The only information available to indicate the languages acceptable to a user are either preferences maintained independently of each HTTP transaction or within the search query itself. First, user-provided preferences are specified either at the Web client or Web server. Client-side preferences, such as languages accepted by a Web browser, are communicated

through request message headers. Server-side preferences are specified via search engine options and are maintained independent of each HTTP transaction using cookies, which must be retrieved from the Web client prior to executing a search, or via a log-in procedure.

Although effective at specifying accepted languages, users seldom explicitly set language preferences in practice. As well, language preferences are often too restrictive, presenting an all-or-nothing paradigm. The language preferences function as a search result filter, providing only those search results in the preferred language and disallowing those search results in related or alternate languages.

Similarly, default settings for specifying accepted languages, either client- or server-side, can further complicate providing suitable search results. Often, default settings can be incorrect. For instance, English could be specified as a default language preference by virtue of a Web browser option, but may be unsuitable for presenting search results to a non-English proficient user.

Second, query-based preferences are derived from the terms in a given search query. Search query terms, however, are not reliable for determining language preferences for several reasons. First, proper nouns, such as the name of a person, place or thing, are often language-independent and are a poor indicator of the language desired for search result presentation. For instance, a search engine will be unable to determine accepted languages for a search query consisting of the proper name "Elvis." Second, search queries, particularly when specified in key words, often consist of only a few individual words, which generally fail to provide sufficient context from which to determine a language preference. Like proper names, individual words can be language-independent or language-misleading. For instance, a search engine could be misled by a search query consisting of the words "Waldorf Astoria."

Accordingly, there is a need to provide an approach to dynamically determining language preferences for the presentation of search results to a user. Preferably, such an approach would accommodate both preferred and lesser preferred languages, which are acceptable to the user, and include both related and alternate languages within the language preferences.

There is a further need for an approach to presenting search results in an ordered fashion in accordance with user preferred languages. Preferably, such an approach would order or score search results to favor those search results in preferred languages while accommodating those search results in other languages.

DISCLOSURE OF THE INVENTION

The present invention provides a system and method for dynamically determining preferred languages and ordering search results in response to a search query. User preferred and

less preferred languages are determined based on an evaluation of search query, user interface, and search result characteristics. Search query characteristics are determined from metadata describing the search query. User interface characteristics are determined also using the search query metadata, as well as client-side and server-side preferences and the Internet protocol (IP) address of the client. Search result characteristics are determined based on an evaluation of each search result. Search results retrieved responsive to the search query are ordered based on the method utilized by the search engine to organize the search results in consideration of the preferred and, if selected, less preferred languages. The search results are ordered by either a predetermined shifting factor or by adjusting a numerical score assigned to each search result by a weighting factor and resorting the search results.

An embodiment provides a system and method for ordering search results. At least one preferred language applicable to search results generated responsive to a search executed on potentially retrievable information and provided in a plurality of search result languages is dynamically determined. At least some of the search results are ordered in consideration of the at least one preferred language.

A further embodiment provides a system and method for providing preferred language ordering of search results. A search query describing potentially retrievable information provided in a plurality of search result languages is received. A search is executed by evaluating the search query against information characteristics maintained in a searchable data repository. At least one preferred language applicable to search results generated responsive to the executed search is dynamically determined. At least some of the search results are ordered in consideration of the at least one preferred language.

A further embodiment provides a system and method for dynamically determining language preferences and ordering of search results. A search query request message is received and at least one of terms and attributes are parsed from the search query request message to identify potentially retrievable Web content provided in a plurality of search result languages. A search is executed by evaluating the at least one of terms and attributes against information characteristics maintained in a searchable data repository and search results are generated responsive to the executed search. At least one preferred language is determined. Characteristics of at least one of the search query request message, a user interface used for sending the search query request message, and the search results are evaluated. The at least one preferred language is selected based on the evaluated characteristics. At least some of the search results are ordered in consideration of the at least one preferred language. The search results are presented as search result response messages.

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.

DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a block diagram showing a system for providing preferred language ordering of search results, in accordance with the present invention.

FIGURE 2 is a functional block diagram showing the search engine of FIGURE 1.

FIGURE 3 is a process flow diagram showing search query execution and search results processing by the search engine of FIGURE 1.

FIGURE 4 is a data structure diagram showing, by way of example, a request message for receipt by the search engine of FIGURE 1.

FIGURE 5 is a data structure diagram showing, by way of example, a response message for dispatch by the search engine of FIGURE 1.

FIGURE 6 is a flow diagram showing a method for providing preferred language ordering of search results, in accordance with the present invention.

FIGURE 7 is a flow diagram showing the routine for determining preferred languages for use in the method of FIGURE 6.

FIGURE 8 is a flow diagram showing the function for evaluating search query characteristics for use in the routine of FIGURE 7.

FIGURE 9 is a flow diagram showing the function for evaluating user interface characteristics for use in the routine of FIGURE 7.

FIGURE 10 is a flow diagram showing the routine for ordering search results for use in the method of FIGURE 6.

FIGURE 11 is a flow diagram showing the routine for ordering search results by a shifting factor for use in the routine of FIGURE 10.

FIGURE 12 is a flow diagram showing the routine for ordering search results by a weighting factor for use in the routine of FIGURE 10.

BEST MODE FOR CARRYING OUT THE INVENTION

FIGURE 1 is a block diagram showing a system 10 for providing preferred language ordering of search results, 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.

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.

For Web content exchange and, in particular, to transact searches, each client 12 executes a Web browser 18 ("browser") through which search queries are sent to a Web server 20 executing on the server 11. Each search query describes or identifies information, generally in the form of Web content, which is potentially retrievable via the Web server 20. 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 search results. 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.

The search results are sent back to the 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 search results. 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 internetwork 13 and which are preferably accessible by each client 12.

The search engine 21 preferably identifies the Web content 22 best matching the search query terms to provide high quality search results, such as described in S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Search Engine" (1998) and in U.S. Patent No. 6,285,999, issued September 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 FIGURE 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.

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.

FIGURE 2 is a functional block diagram showing the search engine 21 of FIGURE 1. Each component is a computer program, procedure or process written as source code in a conventional programming language, such as the C++ programming language, and is presented for execution by one or more CPUs as object or byte code in a uniprocessing, distributed or parallelized configuration, as is known in the art. The various implementations of the source code and object and byte codes can be held on a computer-readable storage medium or embodied on a transmission medium in a carrier wave.

The search engine 21 consists of five components: parser 31, indexer 32, scorer 33, language promoter 34, and presenter 35. Briefly, the search engine 21 receives a search query 36 communicated via a browser 18 from a user 19, executes a search, generates search results 38, orders the search results 38 in consideration of language preferences, and sends the ordered search results 37. The search query 36 is preferably provided as a HTTP-compliant request message and the ordered search results 37 are preferably provided as HTTP-compliant response messages, as further described below respectively with reference to FIGURES 4 and 5, although other forms of request and response exchanges are feasible, as would be recognized by one skilled in the art.

In more detail, the parser 31 receives the search query 36. Each search query 36 describes potentially retrievable information, such as Web content 22. The parser 31 then parses the search query 36 into individual tokens. The tokens include header values constituting metadata 48, and an entity body containing the actual search query. The metadata 48 is copied to the language promoter 34.

The indexer 32 executes the search by evaluating the search query 36 against information characteristics maintained in a searchable data repository 49. The information characteristics are either the actual Web content 22 or metadata, such as hyperlinks, describing terms and attributes used to identify Web content. Other structures and organizations of a searchable data repository 49 are feasible, as would be recognized by one skilled in the art. Upon completing the search, the indexer 32 generates a set of search results 38 by applying the characteristics specified in the search query 36 to the stored information. Other structures and organizations of a searchable data repository 49 are feasible, as would be recognized by one skilled in the art.

Potentially, the indexer 32 can identify thousands or even millions of search results 38, so only a subset of the search results 38, typically between 100 to 10,000, are retained as the most promising search results 38. Targeted search results (not shown) can also be introduced, such as advertising or topical information content. The most promising search results 38 are then qualitatively ranked or scored by degree of match to the search query terms. The search results 38 can be numerically scored to reflect a relative quality or goodness of match. The scorer 33 assigns a numerical score 41 to each search result 38 for indicating a quality of match.

The language promoter 34 performs two primary functions. First, the language promoter 34 determines one or more preferred languages 39 and, optionally, one or more less preferred languages 40 for each search query 36, as further described below with reference to FIGURE 7. In one embodiment, such preferred languages 39 may be determined using a language selector (not shown). Second, the language promoter 34 orders the search results 38 in consideration of the preferred languages 39 and, if available, the less preferred languages 40, as further described below with reference to FIGURE 10. In one embodiment, such search results 38 may be ordered using a search result orderer (not shown). For efficiency, the language promoter 34 preferably orders a subset of the most promising search results 38, typically in the range of 15 to 30 search results, although other ranges could be used, as would be recognized by one skilled in the art.

In the described embodiment, the scorer 33 assigns those search results 38 having a higher degree of match a commensurately higher rank relative to other search results 38. For instance, if Spanish was a preferred language 39, those search results 38 in Spanish would have a higher degree of match than search results 38 in, say, English. However, those search results 38 in a less preferred language, such as Portuguese, could also have a higher degree of match than search results 38 in English, but lower degree of match than search results 38 in Spanish. Alternatively, a counter ranking approach could be used whereby the scorer 33 assigns those search results 38 having a higher degree of match a commensurately lower rank relative to other

search results 38. Other styles, assignments or definitions of search result ranking are feasible, as would be recognized by one skilled in the art.

In another embodiment, the search results 38 are numerically scored to reflect a relative quality or goodness of match. The scorer 33 assigns a numerical score 41 to each search result 38 for indicating a relative quality of match, with higher numerical scores 41 to reflect better quality than lower numerical scores 41. For instance, if Spanish was a preferred language 39, those search results 38 in Spanish would have a higher numerical score 41 than search results 38 in, say, English. However, those search results 38 in a less preferred language, such as Portuguese, could also have a higher numerical score 41 than search results 38 in English, but lower numerical score 41 than search results 38 in Spanish. Alternatively, a counter scoring approach could be used whereby the scorer 33 assigns lower numerical scores 41 to reflect better quality than higher numerical scores 41. Other styles, assignments or definitions of search result scoring are feasible, as would be recognized by one skilled in the art.

To determine the preferred languages 39 and less preferred languages 40, the language promoter 34 evaluates search query characteristics (SQ Chars) 43, user interface characteristics (UI Chars) 44, and search result characteristics (SR Chars) 45. The search query characteristics 43 may be determined from the metadata 48. The user interface characteristics 44 may be determined from the metadata 48, and any available language preferences 42, which may be maintained by the server 11 (server-side) independently of each search query 36. The search result characteristics 45 may be determined from the search results 38.

The language promoter 34 orders the search results 38. In one embodiment, non-numerically ordered search results 38 are ordered by a shifting factor 46, as further describe below with reference to FIGURE 11. In the described embodiment, a constant shifting factor 46 of two (2.0) is employed to demote search results 38 in non-preferred languages by two positions. This shifting factor 38 is suitable when search results 38 are ordered by decreasing degree of match to the search query 36. For example, a search result 38 in a non-preferred language occurring in the third position of a list of the search results 38 would be demoted to the fifth position. Other forms of shifting factors could be employed as well. For instance, the shifting factor 46 could promote search results 38 in non-preferred languages. As well, the shifting factor 46 could order the search results 38 by mathematical function, using, for example, additive, subtractive, multiplicative, fractional, divisional, and logarithmic factors, or may be a formula or function, as well as various combinations and arrangements thereof, as would be recognized by one skilled in the art.

In another embodiment, preferable when an adjustment formula is available for a range of numerical scores 41, numerically scored search results 38 are ordered by a weighting factor 47, as further describe below with reference to FIGURE 12. Such weighting factor 47 may be a constant, mathematical function using, for example, additive, subtractive, multiplicative, fractional, divisional, and logarithmic factors, or may be a formula or function, as well as various combinations and arrangements thereof, as would be recognized by one skilled in the art. In the described embodiment, the numerical scores 41 increase with the quality of match. For example, a higher numerical score 41 reflects a better quality than lower numerical scores 41. Alternatively, a system may be employed whereby the numerical scores 41 could decrease with quality of match, as would be recognized by one skilled in the art. In the described embodiment, two weighting factors 47 are used to increase the numerical score 41 of each search result 38 depending upon whether the search result 38 is in one of the preferred languages 39 or the less preferred languages 40. For search results 38 in one of the preferred languages 39, a weighting factor w_{LP} is provided by the equation:

$$w_{LP} = \left\{ \forall s_{1 \to n} : s_i = \frac{s_i + 1}{2} \right\}$$

For search results 38 in one of the less preferred languages 40, a weighting factor w_{LPL} is provided by the equation:

$$w_{LPL} = \left\{ \forall s_{1 \to n} : s_i = \frac{(s_i \times 2) + 1}{3} \right\}$$

These weighting factors are most suitable when numerical scores 41 range between 0.0 and 1.0 and are approximately uniformly distributed. In the described embodiment, the search results 38 having a given score s_i are promoted more when associated with one of the preferred languages 39 than when associated with one of the less preferred languages 40. However, the search results 38 associated with one of the less preferred languages 40 could instead be promoted more than the search results 38 associated with one of the preferred languages 39, as would be recognized by one skilled in the art. Alternatively, weighting factors 47 could be used to decrease the numerical score 41 of each search result 38 depending upon whether the search result 38 is in one of the preferred languages 39 or the less preferred languages 40. Other ranges of numerical scores and forms of weighting factors could be employed as well, as would be recognized by one skilled in the art.

In a further embodiment, the shifting factor 46 and the weighting factor 47 can be adjusted to accommodate less or more reliable dynamic preferred language determination. For instance, a short search query 36 or sparse search results 37 might lower the accuracy of the

dynamic preferred language determination due to less context with which to work. The shifting factor 46 and the weighting factor 47 would be relaxed to less aggressively order the search results 37 Alternatively, a long search query 36 or lengthy search results 37 might increase the accuracy and the shifting factor 46 and the weighting factor 47 would be increased to more aggressively order the search results 37.

Finally, the presenter 35 presents the ordered search results 37 to the user 19 via the browser 18. Typically, only a part of the ordered search results 37 need be presented since the full set of ordered search results 37 can exceed available presentation space on the browser 18. Presentation is the communication of the ordered search results 37 by means of a search result response message.

FIGURE 3 is a process flow diagram 50 showing search query execution and search results processing by the search engine 21 of FIGURE 1. The process flow proceeds in five primary phases. First, a search query 52 is accepted and parsed (process 51) and is forwarded to the next phase. A search is executed (process 53) against a stored data repository and search results 54 are forwarded to the next phase. The preferred and less preferred languages 56 of the user 19 are determined (process 55) based on metadata and available context. The preferred and less preferred languages 56 are forwarded to the next phase for use in ordering the search results 54 (process 57). In the final phase, the ordered search results 58 are forwarded and presented (process 59). The phases of determining the preferred and less preferred languages (process 55) and ordering the search results (process 57) enhance the quality of the search results by tailoring the search results in accordance with dynamically determined user language preferences.

FIGURE 4 is a data structure diagram showing, by way of example, a request message 60 for receipt by the search engine 21 of FIGURE 1. The request message 60 is an HTTP-compliant request message, such as described in D. Gourley and E. Totty, "HTTP, the Definitive Guide," Ch. 3, pp. 43-73, O'Reilly and Assocs., Sebastopol, CA (2002), the disclosure of which is incorporated by reference. The request message 60 consists of three parts: start line 61, headers 62, and entity body 63. The start line 61 identifies an HTTP method, such as, "POST," which sends input data from the browser 18 to the search engine 21. The start line 61 also includes a request Uniform Resource Locator (URL) 68 and HTTP version identifier. The exemplary request URL, "/search.cgi," identifies a search request.

The headers 62 consist of zero or more MIME-compliant name and value pairings, which provide the metadata 48 describing the characteristics of the interface of the browser 18 and the entity body 63, that is, the search query itself. Four MIME-compliant headers provide metadata 48 instrumental in determining those languages acceptable to a requesting user, as follows:

(1) *Accept_Charset* (64): Indicates the character sets that are acceptable or preferred by the requesting browser 18, for instance, *ISO-Latin-1*, an eight-bit extension to ASCII that supports Western European languages. *ISO-Latin-1* is also known as *ISO-8859-1*.

(2) *Accept_Language* (65): Indicates the languages that are acceptable or preferred by the user 19, in order of preference and optionally including a quality factor *q*, for instance, strong preference for French *(fr)* and lesser preference for English *(en)*.

(3) *Content_Type* (66): Describes the media type of the entity body 63, for instance, plain text *(text/plain)* and can indicate the character sets used to encode the entity body 63, for instance, *ISO-Latin-1*, as a parameter.

(4) *Content_Language* (67): Indicates the natural language in which the entity body 63 is expressed, for instance, French *(fr)*.

Other forms of request message formats using equivalent or related protocols and providing similar information as the forgoing headers 62, as well as other headers and parameters, are feasible, as would be recognized by one skilled in the art.

FIGURE 5 is a data structure diagram showing, by way of example, a response message 70 for dispatch by the search engine 21 of FIGURE 1. The response message 70 is an HTTP-compliant response message, such as described in D. Gourley and E. Totty, *Id.*, the disclosure of which is incorporated by reference. The response message 70 also consists of three parts: start line 71, headers 72, and entity body 73. The start line 71 includes an HTTP identifier, response status code, and human-readable reason phrase.

The headers 72 consist of zero or more MIME-compliant name and value pairings, which provide the metadata describing the characteristics of the entity body 73, that is, each search result. Two MIME-compliant headers provide metadata instrumental in determining those languages acceptable to a requesting user, as follows:

(1) *Content_Type* (74): Describes the media type of the entity body 73, for instance, plain text *(text/plain)* and indicates the character sets used to encode the entity body 73, for instance, *ISO-Latin-1*, as a parameter.

(2) *Content_Language* (75): Indicates the natural language in which the entity body 73 is expressed, for instance, English *(en)*.

Other forms of response message formats using equivalent or related protocols and providing similar information as the forgoing headers 72, as well as other headers and parameters, are feasible, as would be recognized by one skilled in the art.

FIGURE 6 is a flow diagram showing a method 80 for providing preferred language ordering of search results 38, in accordance with the present invention. The method 80 is

described as a sequence of process operations or steps, which can be executed, for instance, by a search engine 21 (shown in FIGURE 1).

A search query 36 is accepted from a user 19 and parsed into metadata 48 (block 81). A search is executed on a searchable data repository 49 by evaluating the search query 36 against information characteristics maintained in the searchable data repository 49 and search results 38 are generated (block 82). Since thousands or even millions of search results 38 can potentially be generated, only a subset of the search results 38, typically between 100 to 10,000, are retained as the most promising search results 38. Targeted search results (not shown) can also be introduced, such as advertising or topical information content. Prior to providing the search results 38 to the browser 18, the search results 38 can be temporarily staged as "raw" Web pages, structured data, or unstructured data, from which metadata describing the characteristics of each search result 38 can be extracted, as is known in the art. The search results 38 are qualitatively ordered by degree of match to the search query 36 (block 83) to provide a ranking or scoring, including a numerical score 41, reflecting search result quality, as described above with reference to FIGURE 2.

Up to this point, the search results 38 have been identified and ranked or scored. The preferred languages 39, as well as the less preferred languages 40, of the requesting user 19 are then determined (block 84), as further described below with reference to FIGURE 7. The search results 38 are then ordered by degree of match to the preferred languages 39 and, if identified, less preferred languages 40 (block 85), as further described below with reference to FIGURE 10. For efficiency, preferably only a subset of the most promising search results 38, typically in the range of 15 to 30 search results, are ordered, although other ranges could be used, as would be recognized by one skilled in the art. Finally, the ordered search results 37 are presented via the browser 18. Typically, only a part of the ordered search results 37 need be presented since the full set of ordered search results 37 can exceed available presentation space on the browser 18. In the described embodiment, the search query 36 is provided as an HTTP-compliant request message 60 and each search result is provided back to the user 19 as an HTTP-compliant response message 70.

In a further embodiment, the preferred languages 39 and less preferred languages 40 are stored for future use while executing search queries 36 from the same requesting user 19. As well, the ability to present the ordered search results 37 using preferred language ordering could be controlled by enabling or disabling presentation in the preferred languages 39 and, if available, the less preferred languages 40, using a "toggle" provided via the user interface. In a still further embodiment, the ordered search results 37 are grouped together in each of the

preferred languages 39 and, if available, the less preferred languages 40, prior to presentation to the requesting user 19. Alternatively, the ordered search results 37 in the preferred languages 39 and, if available, the less preferred languages 40 can be arranged for presentation next to those search results in non-preferred languages, such as by using adjacent columns or cells in a table. The routine then terminates.

FIGURE 7 is a flow diagram showing the routine 90 for determining preferred languages 39 for use in the method 80 of FIGURE 6. The purpose of this routine is to determine the preferred languages 39 and any less preferred languages 40 of the requesting user 19 based on search query characteristics 43, user interface characteristics 44, and search result characteristics 45.

First, the search query characteristics 43 are evaluated (block 91) based on the metadata 48 parsed from the search query 36, as further described below with reference to FIGURE 8. If specific languages can be determined based on the search query characteristics 43 (block 92), the specific languages are selected as the preferred languages 39 (block 93).

Next, if no specific languages can be determined from the search query characteristics 43 (block 92), the user interface characteristics 44 are evaluated (block 94), as further described below with reference to FIGURE 9. If the user interface does not define English as an accepted language (block 95), each non-English language is selected as a preferred language 39 and English is selected as a less preferred language 40 (block 96).

Next, if English is provided as the accepted language of the user interface (block 95), the search results characteristics 45 are evaluated (block 97). In the described embodiment, the search results 38 are provided in one of two formats. First, the search results 38 can be grouped as a collection of "raw" Web pages from which language characteristics can be determined. Second, the search results 38 can be organized into metadata describing the various characteristics, including language characteristics, and content of the Web pages corresponding to the search results 38. The predominant language of each search result 38 can be dynamically determined through content analysis, such as described in U.S. Patent No. 6,167,369, issued December 26, 2000 to Schulze, the disclosure of which is incorporated by reference. If a majority of the search results 38 are in English (block 98), English is selected as a preferred language 39 (block 99). Otherwise, no preferred languages 39 or less preferred languages 40 are assumed (block 100) and the routine returns.

For each instance in which one or more preferred language 39 has been selected (blocks 93, 96 and 99), related languages, if any, can optionally be selected as additional preferred languages 39 (block 101). Related language include those languages and dialects sharing a

common basis whereby users proficient in one such language are able to comprehend, perhaps with only slight difficulty, related languages. For instance, a user proficient in Spanish can often comprehend information provided in Portuguese. Following any additions to the preferred languages 39, the routine returns.

FIGURE 8 is a flow diagram showing the function 110 for evaluating search query characteristics 43 for use in the routine 90 of FIGURE 7. The purpose of this function is to determine any preferred languages 39 based on available metadata 48 parsed from the headers 62 of the search query request message 60.

First, any available metadata 48 corresponding to the headers 62 of the search query request message 60 are retrieved (block 111). The languages of the content provided in the entity body 63 are determined (block 112) by evaluating the parameters of the *Content_Language* header 67. If specified (block 113), the specified languages are returned (block 114). Otherwise, the character sets of the content provided in the entity body 63 are determined (block 115) by evaluating the *Content_Type* header 66. If specified as language-specific character sets (block 116), the specified languages are returned (block 117). For example, the *ISO-2022-JP* is a Japanese-specific character set. Otherwise, no language preferences are returned (block 118).

FIGURE 9 is a flow diagram showing the function 120 for evaluating user interface characteristics 44 for use in the routine 90 of FIGURE 7. The purpose of this function is to determine any preferred languages 40 based on user interface characteristics.

The user interface is generated by the search engine 21 and express language preferences can often be specified as user options. Thus, any available language preferences are first retrieved (block 122). Server-side language preferences are maintained either directly in a cookie stored at the client, or by the search engine 21 and accessed using cookies or log-in procedures to uniquely identify each requesting user 19. If available, the appropriate language-identifying cookie is retrieved, or the appropriate user log-in procedure is performed. The stored language preference are retrieved and used to determine any server-side language preferences (block 122). If specified (block 123), the specified languages are returned (block 124).

Browsers 18 can limit the languages in which search results 38 are accepted as client-side language preferences. Thus, any available metadata 48 corresponding to the headers 62 of a search query request message 60 are retrieved (block 125) and the languages accepted by the user interface of the browser 18 are determined (block 126) by evaluating the *Accept_Language* header 65. In addition to specifying accepted languages, the *Accept_Language* header 65 can include a quality factor *q* that indicates a degree of language preference on a scale of 0.0 through

1.0. If specified (block 127), the specified languages are returned (block 120) with those languages having a quality factor *q* less than 1.0 being specified as less preferred languages 40.

If no accepted languages for the user interface of the browser 18 are specified (block 127), the character sets accepted by the user interface of the browser 18 are determined (block 128) by evaluating the *Accept_Charset* header 64. If specified as language-specific character sets (block 130), the specified languages are returned (block 131).

Otherwise, no language preferences can be determined by either server-side or client-side language preferences. However, a language preference might still be determinable based on information available in the URL 68 and the client location. The location of the IP domain identified in the URL 68 is determined (block 132). For example, a URL 68 specified as "*www.acme.at*" has an IP domain of "*.at*," which indicates an Austrian IP domain. However, certain IP domains, such as "*.com*," are so widely-used that no useful language preference inferences can be drawn. If the URL 68 provides a useful IP domain (block 133), the languages of the country to which the IP domain is assigned is determined (block 134) and returned (135). If not useful (block 133), the location of the client 12 from which the search query 36 was sent is determined (block 136) by evaluating a *Client_IP* parameter, which can be parsed from the header of the Transmission Control Protocol (TCP) packet within which the search query 36 was sent. The *Client_IP* parameter provides an IP address, which is a 32-bit numeric address written as four numbers separated by periods. An IP domain can be determined by a Domain Name Service lookup of the last number in the IP address. Again, if useful in inferring a language preference (block 137), the languages of the IP domain are determined (block 138) and returned (block 139). In the described embodiment, the predominant language for the domain specified in the IP address of the client 12 is selected. Otherwise, if no IP address is provided (block 137), English is returned (block 140).

FIGURE 10 is a flow diagram showing the routine 145 for ordering search results 38 for use in the method 80 of FIGURE 6. The purpose of this routine is to order the search results 38 based on the ranking or ordering method used, if any, by the search engine 21 in consideration of the preferred languages 39 and, if available, the less preferred languages 40.

First, the approach utilized by the search engine 21 to rank or order the search results 38 is determined (block 146). If the numerical scores 41 assigned to the search results 38 are suitable for numerical adjustment (block 147), a numerical scoring approach is utilized, whereby the search results 38 are ordered using a weighting factor 47 (block 149), as further described below with reference to FIGURE 12. Otherwise, the search results 38 are ordered by using a

shifting factor 26 (block 148), as further described below with reference to FIGURE 11. The routine then returns.

FIGURE 11 is a flow diagram showing the routine 150 for ordering search results 38 by a shifting factor 46 for use in the routine 130 of FIGURE 9. The purpose of this routine is to shift search results 38 in a non-preferred language down in relation to search results 38 in a preferred language 39 or less preferred language 40. In the described embodiment, a constant shifting factor 46 of two (2.0) is employed to demote search results 38 in non-preferred languages by two positions. This shifting factor 38 is suitable when search results 38 are ordered by decreasing degree of match to the search query 36. For example, a search result 38 in a non-preferred language occurring in the third position of a list of the search results 38 would be demoted to the fifth position. Other forms of shifting factors could be employed as well. For instance, the shifting factor 46 could promote search results 38 in non-preferred languages. As well, the shifting factor 46 could order the search results 38 by mathematical functions, including additive, subtractive, multiplicative, fractional, divisional, and logarithmic factors, as well as various combinations and arrangements thereof, as would be recognized by one skilled in the art.

On one embodiment, the routine operates on a subset less than or equal to the total number of search results 38, although other forms of subset selection criteria could be used, as would be recognized by one skilled in the art. A variable *n* is set to the number of search results to be displayed to the user 19, while an upper limit *UL* for preferred language 39 and an upper limit for less preferred languages *LPUL* are both set to the lesser of the number of search results 38, *n*, and twice *n* (block 151). The search results 38 are then ordered in an iterative processing loop (blocks 153-166) as follows. First, an index *j* is set to the upper limit *UL* (block 152) and processing is performed while the index *j* is positive (block 153). The search results are maintained in an array *Result[]*. If *Result [j]* is not in a preferred language *Pref_Lang* and *Result [j]* is not in a less preferred language *Less_Pref_Lang* (block 154), *Result [j]* is demoted by the shifting factor 46, as follows. A variable *target_pos* is set to the lesser of twice *j* and upper limit *UL* (block 155) and a temporary variable, *temp*, is set to *Result [j]* (block 156). The remaining search results 38 are promoted by shifting *Results [j+1]* through *Results [target_pos]* up by one (block 157) and *Result [target_pos]* is set to *temp* (block 158). The upper limit *UL* is set to *target_pos* minus one (block 159) and the less preferred upper limit *LPUL* is set to the lesser of the less preferred upper limit *LPUL* and upper limit *UL* (block 160).

If *Result [j]* is in a less preferred language *Less_Pref_Lang* (block 161), *Result [j]* is demoted by the shifting factor 46, as follows. The variable *target_pos* is set to the lesser of the floor of 1.5 times *j* and the less preferred upper limit *LPUL* (block 162) and a temporary

variable, *temp*, is set to *Result [j]* (block 163). The remaining search results 38 are shifted by promoted *Results [j+1]* through *Results [target_pos]* up by one (block 164) and *Result [target_pos]* is set to *temp* (block 165). The less preferred upper limit (*LPUL*) is set to *target_pos* minus one (block 166).

Finally, the index *j* is set to *j* minus one (block 167) and processing continues with the next loop iteration (block 168), after which the routine returns.

FIGURE 12 is a flow diagram showing the routine 170 for ordering search results 38 by a weighting factor 47 for use in the routine 130 of FIGURE 9. The purpose of this routine is to order the search results 38 by recalculating the numerical score 41 assigned to each search result 38 to favor those search results in either one of the preferred languages 39 or less preferred languages 40. On one embodiment, the routine operates on a subset less than or equal to the number of search results 38, which equals the number of search results 38 to be displayed to the user multiplied by a margin, for instance, 2 or 3, although other forms of subset selection criteria could be used, as would be recognized by one skilled in the art.

In the described embodiment, the numerical scores 41 increase with the quality of match. For example, a higher numerical score 41 reflects a better quality than lower numerical scores 41. Alternatively, a system may be employed whereby the numerical scores 41 could decrease with quality of match, as would be recognized by one skilled in the art. In the described embodiment, two weighting factors 47 are used to increase the numerical score 41 of each search result 38 depending upon whether the search result 38 is in one of the preferred languages 39 or the less preferred languages 40. For search results 38 in one of the preferred languages 39, a weighting factor w_{LP} is provided by the equation (1):

$$w_{LP} = \left\{ \forall s_{1 \to n} : s_i = \frac{s_i + 1}{2} \right\} \qquad (1)$$

For search results 38 in one of the less preferred languages 40, a weighting factor w_{LPL} is provided by the equation (2):

$$w_{LPL} = \left\{ \forall s_{1 \to n} : s_i = \frac{(s_i \times 2) + 1}{3} \right\} \qquad (2)$$

These weighting factors are most suitable when numerical scores 41 range between 0.0 and 1.0 and are approximately uniformly distributed. In the described embodiment, the search results 38 having a given score s_i are promoted more when associated with one of the preferred languages 39 than when associated with one of the less preferred languages 40. However, the search results 38 associated with one of the less preferred languages 40 could instead be promoted more than the search results 38 associated with one of the preferred languages 39, as would be recognized

by one skilled in the art. Alternatively, weighting factors 47 could be used to decrease the numerical score 41 of each search result 38 depending upon whether the search result 38 is in one of the preferred languages 39 or the less preferred languages 40. Other ranges of numerical scores and forms of weighting factors could be employed as well, as would be recognized by one skilled in the art.

The routine operates on a subset less than or equal to the total number of search results 38 and recalculates the numerical scores 41 through an iterative processing loop (blocks 172-177) as follows. A variable *n* is set to the number of search results to be multiplied by a margin, for instance 2 or 3 (block 171). The search results 38 are maintained in an array *Results[]*. The numerical score 41 for each search result 38 is recalculated in the iterative processing loop (blocks 172-177) indexed by a variable *i*. During each iteration (block 172), if *Result [i]* is in a preferred language *Pref_Lang* (block 173), *Score [i]* is set to half the quantity *Score[i]* plus one (block 167), that is, Equation (1). Otherwise, if *Result [i]* is in a less preferred language *Less_Pref_Lang* (block 175), *Score[i]* is set to one-third the quantity two times *Score[i]* plus one (block 176) , that is, Equation (2). Otherwise, no numerical score adjustment is required. Processing continues with the next iteration (block 177). After all iterations, the search results 38 are resorted (block 178), after which the routine returns.

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.

CLAIMS:

1. A system (10) for ordering search results (38), comprising:

a language selector (21) dynamically determining (84) at least one preferred language (39) applicable to search results (38) generated responsive to a search (36) executed on potentially retrievable information (22) and provided in a plurality of search result languages (37); and

a search result orderer (34) ordering at least some of the search results (38) in consideration of the at least one preferred language (39).

2. A system (10) according to Claim 1, further comprising:

a search query characterizer (43) determining (84) the at least one preferred language (39) using a search query (36) specifying the search (36).

3. A system (10) according to Claim 2, further comprising:

the search query characterizer (43) evaluating at least one of at least one language (42) used in the search query (36) and at least one character encoding (36) used in the search query (36).

4. A system (10) according to Claim 1, further comprising:

a user interface characterizer (44) determining (84) the at least one preferred language (39) using user interface characteristics (44).

5. A system (10) according to Claim 4, further comprising:

the user interface characterizer (44) evaluating at least one of at least one predetermined language (42) specified as a stored preference (39) independent of a search query (36) specifying the search (36), at least one language (42) accepted by the user interface (44), at least one character encoding (36) accepted by the user interface (44), and a network address (36) of a client application (36) from which the search query (36) was submitted.

6. A system (10) according to Claim 1, further comprising:

a search result characterizer (45) determining (84) the at least one preferred language (39) using at least some of the search results (38).

7. A system (10) according to Claim 6, further comprising:

the search result characterizer (45) evaluating at least one language (39) used in one or more of the search results (38).

8. A system (10) according to Claim 1, further comprising:

an indexer (32) ranking the search results (38); and

the search result orderer (34) ordering at least some of the search results (38) relative to the at least one preferred language (39).

9. A system (10) according to Claim 8, further comprising at least one of:

the search result orderer (34) demoting each such search result (38) in a language (40) other than the at least one preferred language (39) and promoting each such search result (38) in the at least one preferred language (39) by at least one position (39); and

the search result orderer (34) promoting each such search result (38) in a language (40) other than the at least one preferred language (39) and demoting each such search result (38) in the at least one preferred language (39) by at least one position (39).

10. A system (10) according to Claim 1, further comprising:

a scorer (33) assigning a numerical score (41) to the search results (38); and

a search result orderer (34) adjusting the numerical score (41) of at least some of the search results (38) in the at least one preferred language (39).

11. A system (10) according to Claim 10, further comprising at least one of:

the search result orderer (34) increasing the numerical score (41) assigned to the search results (38) in the at least one preferred language (39) and maintaining the numerical score (41) assigned to the search results (38) in a language (40) other than the at least one preferred language (39); and

the search result orderer (34) decreasing the numerical score (41) assigned to the search results (38) in the at least one preferred language (39) and maintaining the numerical score (41) assigned to the search results (38) in a language (40) other than the at least one preferred language (39).

12. A system (10) according to Claim 1, further comprising:
the search result orderer (34) sorting the at least some of the search results (38) with adjusted numerical scores (41).

13. A system (10) according to Claim 1, further comprising:
the language selector (21) selecting at least one less preferred language (40).

14. A method (80) for ordering search results (38), comprising:
dynamically determining (84) at least one preferred language (39) applicable to search results (38) generated (82) responsive to a search (36) executed (82) on potentially retrievable information (22) and provided (86) in a plurality of search result languages (37); and
ordering (83) at least some of the search results (37) in consideration of the at least one preferred language (39).

15. A method (80) according to Claim 14, further comprising:
determining (84) the at least one preferred language (39) using a search query (36) specifying the search (36).

16. A method (80) according to Claim 15, further comprising:
evaluating at least one of at least one language (42) used in the search query (36) and at least one character encoding (36) used in the search query (36) .

17. A method (80) according to Claim 14, further comprising:
determining (84) the at least one preferred language (39) using user interface characteristics (44).

18. A method (80) according to Claim 17, further comprising:
evaluating (84) at least one of at least one predetermined language (42) specified as a stored preference (39) independent of a search query (36) specifying the search (36), at least one language (42) accepted by the user interface, at least one character encoding (36) accepted by the user interface, and a network address (36) of a client application (36) from which the search query (36) was submitted.

19. A method (80) according to Claim 14, further comprising:

determining (84) the at least one preferred language (39) using at least some of the search results (38).

20. A method (80) according to Claim 19, further comprising:

evaluating at least one language (42) used in one or more of the search results (38).

21. A method (80) according to Claim 14, further comprising:

ranking (85) the search results (38); and

ordering (85) at least some of the search results (36) relative to the at least one preferred language (39).

22. A method (80) according to Claim 21, further comprising at least one of:

demoting (85) each such search result (38) in a language (40) other than the at least one preferred language (39) and promoting (85) each such search result (38) in the at least one preferred language (39) by at least one position; and

promoting (85) each such search result (38) in a language (40) other than the at least one preferred language (39) and demoting each such search result (38) in the at least one preferred language (39) by at least one position.

23. A method (80) according to Claim 14, further comprising:

assigning (83) a numerical score (41) to the search results (38); and

adjusting (83) the numerical score (41) of at least some of the search results (38) in the at least one preferred language (39).

24. A method (80) according to Claim 23, further comprising at least one of:

increasing (83) the numerical score (41) assigned to the search results (38) in the at least one preferred language (39) and maintaining (83) the numerical score (41) assigned to the search results (38) in a language (42) other than the at least one preferred language (39); and

decreasing (83) the numerical score (41) assigned to the search results (38) in the at least one preferred language (39) and maintaining (83) the numerical score (41) assigned to the search results (38) in a language (42) other than the at least one preferred language (39)

25. A method (80) according to Claim 14, further comprising:

sorting (83) the at least some of the search results (38) with adjusted numerical scores (41).

26. A method (80) according to Claim 14, further comprising:

selecting (84) at least one less preferred language (40).

27. A computer-readable storage medium (30) holding code for performing the method (80) according to Claim 14.

28. A system (10) for providing preferred language (39) ordering of search results (38), comprising:

a parser (31) receiving a search query (36) describing potentially retrievable information (22) provided in a plurality of search result languages (37);

an indexer (32) executing a search (36) by evaluating the search query (36) against information characteristics (49) maintained in a searchable data repository (49); and

a language promoter (34) dynamically determining (84) at least one preferred language (39) applicable to search results (38) generated responsive to the executed search (36), and ordering at least some of the search results (38) in consideration of the at least one preferred language (39).

29. A system (10) according to Claim 28, further comprising:

a search query characterizer (43) determining (84) at least one of a character encoding (36) and a language (42) used in the search query (36) ; and

a language selector (21) selecting at least one language (42) corresponding to the at least one of a character encoding (36) and a language (42) as the at least one preferred language (39).

30. A system (10) according to Claim 28, further comprising:

a user interface characterizer (44) determining (84) at least one language (42) used by a user interface (44) for transacting the search query (36) ; and

a language selector (21) selecting the at least one language (42) used by the user interface (44) as the at least one preferred language (39).

31. A system (10) according to Claim 30, further comprising:

the user interface characterizer (44) determining (84) the at least one language (42) used by the user interface (44) based on at least one of the network address (36) of a client application (36) from which the search query (36) was submitted, at least one of an accepted language (42) and a character set (49) of the client application (36), and a predetermined language (42) specified as a stored preference (39) independent of the search query (36) .

32. A system (10) according to Claim 28, further comprising:

a user interface characterizer (44) determining (84) a default language (42) used by a user interface (44) for transacting the search query (36) ; and

a language selector (21) selecting a language (42) other than the default language (39) as the at least one preferred language (39), and selecting the default language (39) as at least one less preferred language (40).

33. A system (10) according to Claim 32, further comprising:

a search results characterizer(21) determining (84) at least one search result language (42) for at least one search result (38); and

a language selector (21) selecting the at least one search result language (42) as the at least one preferred language (39).

34. A system (10) according to Claim 33, wherein the at least one search result language (42) is used in a majority of the search results (38).

35. A system (10) according to Claim 28, further comprising:

a search result orderer (34) ordering the search results (38) based on a match of a language (42) of such search results (38) to the at least one preferred language (39).

36. A system (10) according to Claim 28, further comprising:

a search result orderer (34) ordering the search results (38) by degree of match of a language (42) of such search results (38) to the at least one preferred language (39).

37. A system (10) according to Claim 36, further comprising:

the search result orderer (34) demoting the search results (38) in a language (42) other than the at least one preferred language (39) by a predefined shifting factor (46).

38. A system (10) according to Claim 37, wherein the predefined factor (46) substantially equals two (2.0).

39. A system (10) according to Claim 36, further comprising:

the search result orderer (34) promoting the search results (38) in a language (42) other than the at least one preferred language (39) by a predefined shifting factor (46).

40. A system (10) according to Claim 28, further comprising:

a scorer (33) ordering each of the search results (38) by degree of match to the information characteristics (49).

41. A system (10) according to Claim 28, wherein the search results (38) are assigned a numerical score (41), further comprising:

a search result orderer (34) increasing the numerical score (41) assigned to at least some of the search results (38) in the at least one preferred language (39).

42. A system (10) according to Claim 41, wherein the numerical score (41) is adjusted in accordance with the formula:

$$s_i = \frac{s_i + 1}{2}$$

where s_i comprises the numerical score (41) for each such search result i (38).

43. A system (10) according to Claim 41, wherein the search results (38) are assigned a numerical score (41), further comprising:

a search result orderer (34) decreasing the numerical score (41) assigned to at least some of the search results (38) in the at least one preferred language (39).

44. A system (10) according to Claim 41, further comprising:

a language selector (21) determining (84) at least one less preferred language (40); and

the search result orderer (34) increasing the numerical score (41) assigned to at least some of the search results (38) in the at least one less preferred language (40).

45. A system (10) according to Claim 44, wherein the numerical score (41) is adjusted in accordance with the formula:

$$s_i = \frac{(s_i \times 2)+1}{3}$$

where s_i comprises the numerical score (41) for each such search result i (38).

46. A system (10) according to Claim 41, further comprising:

a language selector (21) determining (84) at least one less preferred language (40); and

the search result orderer (34) decreasing the numerical score (41) assigned to at least some of the search results (38) in the at least one less preferred language (40).

47. A system (10) according to Claim 28, further comprising:

a presenter presenting the search results (38).

48. A system (10) according to Claim 47, further comprising:

the presenter performing at least one of controlling enablement of presentation of at least some of the search results (38) for each of the at least one preferred language (39), grouping together at least some of the search results (38) for each of the at least one preferred language (39) and arranging at least some of the search results (38) for each of the at least one preferred language (39) next to at least some of the search results (38) for at least one language (42) other than the at least one preferred languages (39) prior to presentation to the user.

49. A system (10) according to Claim 28, further comprising:
a language selector (21) selecting at least one less preferred language (40).

50. A system (10) according to Claim 28, further comprising:
a language selector (21) including one or more related languages (39) in the at least one preferred language (39).

51. A method (80) for providing preferred language ordering (85) of search results (37), comprising:
receiving (81) a search query (36) describing potentially retrievable information (22) provided in a plurality of search result languages (37);
executing (82) a search (36) by evaluating the search query (36) against information characteristics (49) maintained in a searchable data repository (49);
dynamically determining (84) at least one preferred language (39) applicable to search results (38) generated (82) responsive to the executed search (36); and
ordering (85) at least some of the search results (38) in consideration of the at least one preferred language (39).

52. A method (80) according to Claim 51, further comprising:
determining (84) at least one of a character encoding (36) and a language (42) used in the search query (36) ; and
selecting (84) at least one language (42) corresponding to the at least one of a character encoding (36) and a language (42) as the at least one preferred language (39).

53. A method (80) according to Claim 51, further comprising:
determining (84) at least one language (42) used by a user interface (44) for transacting (84) the search query (36) ; and
selecting (84) the at least one language (42) used by the user interface (44) as the at least one preferred language (39).

54. A method (80) according to Claim 53, further comprising:

determining (84) the at least one language (42) used by the user interface (44) based on at least one of the network address (36) of a client application (36) from which the search query (36) was submitted, at least one of an accepted language (37) and a character set (49) of the client application (36), and a predetermined language (42) specified as a stored preference (39) independent of the search query (36) .

55. A method (80) according to Claim 51, further comprising:

determining (84) a default language (39) used by a user interface (44) for transacting the search query (36) ; and

selecting (84) a language (39) other than the default language (39) as the at least one preferred language (39); and

selecting (84) the default language (39) as at least one less preferred language (40).

56. A method (80) according to Claim 51, further comprising:

determining (84) at least one search result language (42) for at least one search result (37); and

selecting (84) the at least one search result language (42) as the at least one preferred language (39).

57. A method (80) according to Claim 56, wherein the at least one search result language (42) is used in a majority of the search results (38).

58. A method (80) according to Claim 51, further comprising:

ordering (85) the search results (38) based on a match of a language (42) of such search results (38) to the at least one preferred language (39).

59. A method (80) according to Claim 51, further comprising:

ordering (85) the search results (38) by degree of match of a language (42) of such search results (38) to the at least one preferred language (39).

60. A method (80) according to Claim 59, further comprising:

demoting (83)the search results (38) in a language (42) other than the at least one preferred language (39) by a predefined shifting factor (46).

61. A method (80) according to Claim 60, wherein the predefined factor substantially equals two (2.0).

62. A method (80) according to Claim 59, further comprising:

promoting (83) the search results (38) in a language (42) other than the at least one preferred language (39) by a predefined shifting factor (46).

63. A method (80) according to Claim 51, further comprising:

ordering (85) each of the search results (38) by degree of match to the information characteristics.

64. A method (80) according to Claim 63, wherein the search results (38) are assigned a numerical score (41), further comprising:

increasing (83) the numerical score (41) assigned to at least some of the search results (38) in the at least one preferred language (39).

65. A method (80) according to Claim 64, wherein the numerical score (41) is adjusted in accordance with the formula:

$$s_i = \frac{s_i + 1}{2}$$

where s_i comprises the numerical score (41) for each such search result i (38).

66. A method (80) according to Claim 64, wherein the search results (38) are assigned a numerical score (41), further comprising:

decreasing (83) the numerical score (41) assigned to at least some of the search results (38) in the at least one preferred language (39).

67. A method (80) according to Claim 64, further comprising:

determining (84) at least one less preferred language (40); and

increasing (84) the numerical score (41) assigned to at least some of the search results (38) in the at least one less preferred language (40).

68. A method (80) according to Claim 67, wherein the numerical score (41) is adjusted in accordance with the formula:

$$s_i = \frac{(s_i \times 2) + 1}{3}$$

where s_i comprises the numerical score (41) for each such search result i (38).

69. A method (80) according to Claim 64, further comprising:

determining (84) at least one less preferred language (40); and

decreasing (84) the numerical score (41) assigned to at least some of the search results (38) in the at least one less preferred language (40).

70. A method (80) according to Claim 51, further comprising:

presenting (86) the search results (38).

71. A method (80) according to Claim 70, further comprising:

performing (86) at least one of controlling enablement (86) of presentation (86) of at least some of the search results (38) for each of the at least one preferred language (39), grouping together at least some of the search results (38) for each of the at least one preferred language (39) and arranging at least some of the search results (38) for each of the at least one preferred language (39) next to at least some of the search results (38) for at least one language (42) other than the at least one preferred languages (39) prior to presentation to the user.

72. A method (80) according to Claim 51, further comprising:

selecting (84) at least one less preferred language (40).

73. A method (80) according to Claim 51, further comprising:

including (84) one or more related languages(39) in the at least one preferred language (39).

74. A computer-readable storage medium (30) holding code for performing the method (80) according to Claim 51.

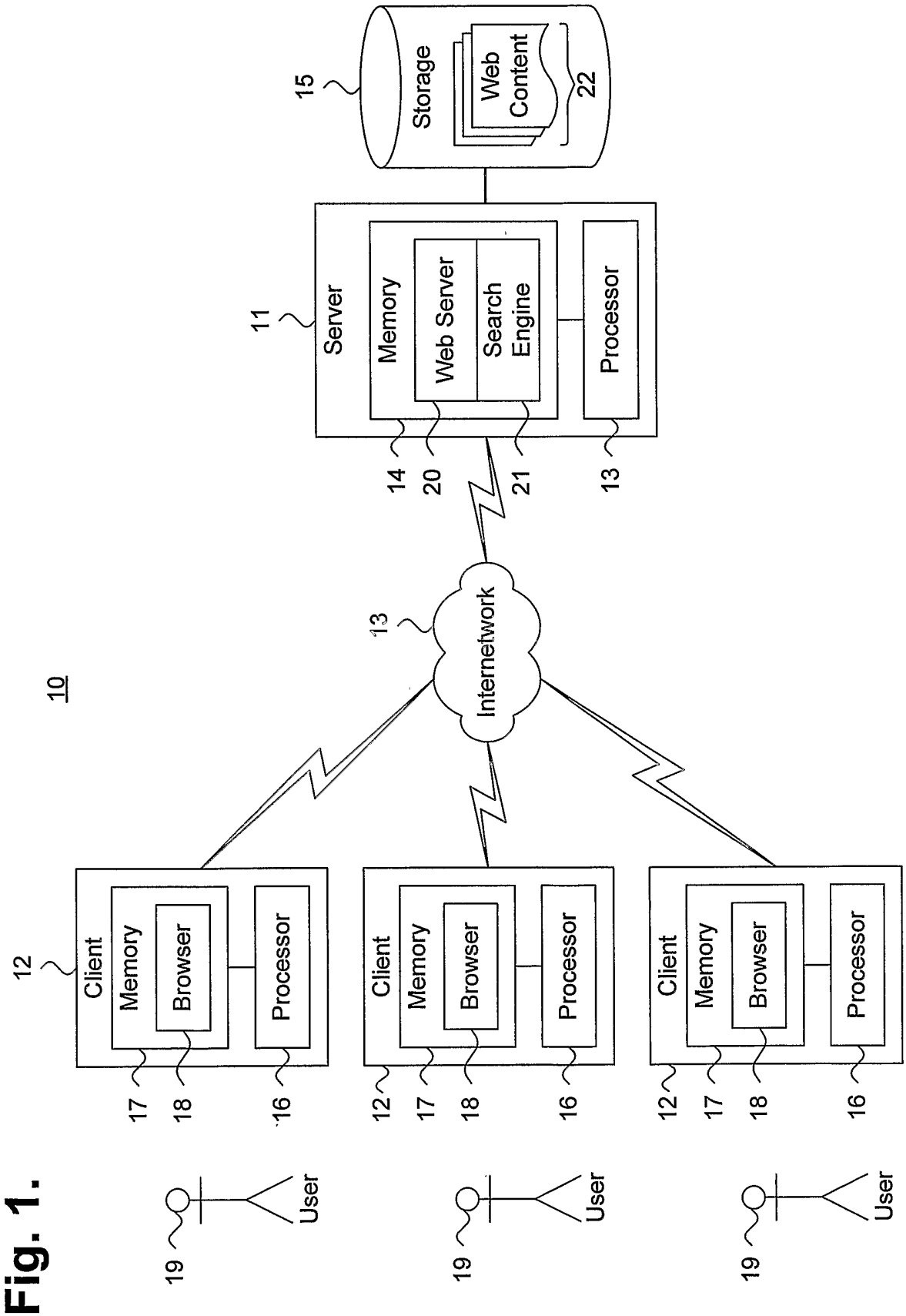

10
11
Server
Memory
Web Server
Search Engine
Processor
14
20
21
13
15
Storage
Web Content
22
13
Internetwork
12
Client
Memory
Browser
Processor
17
18
16
19
User

Fig. 1.

Fig. 2.

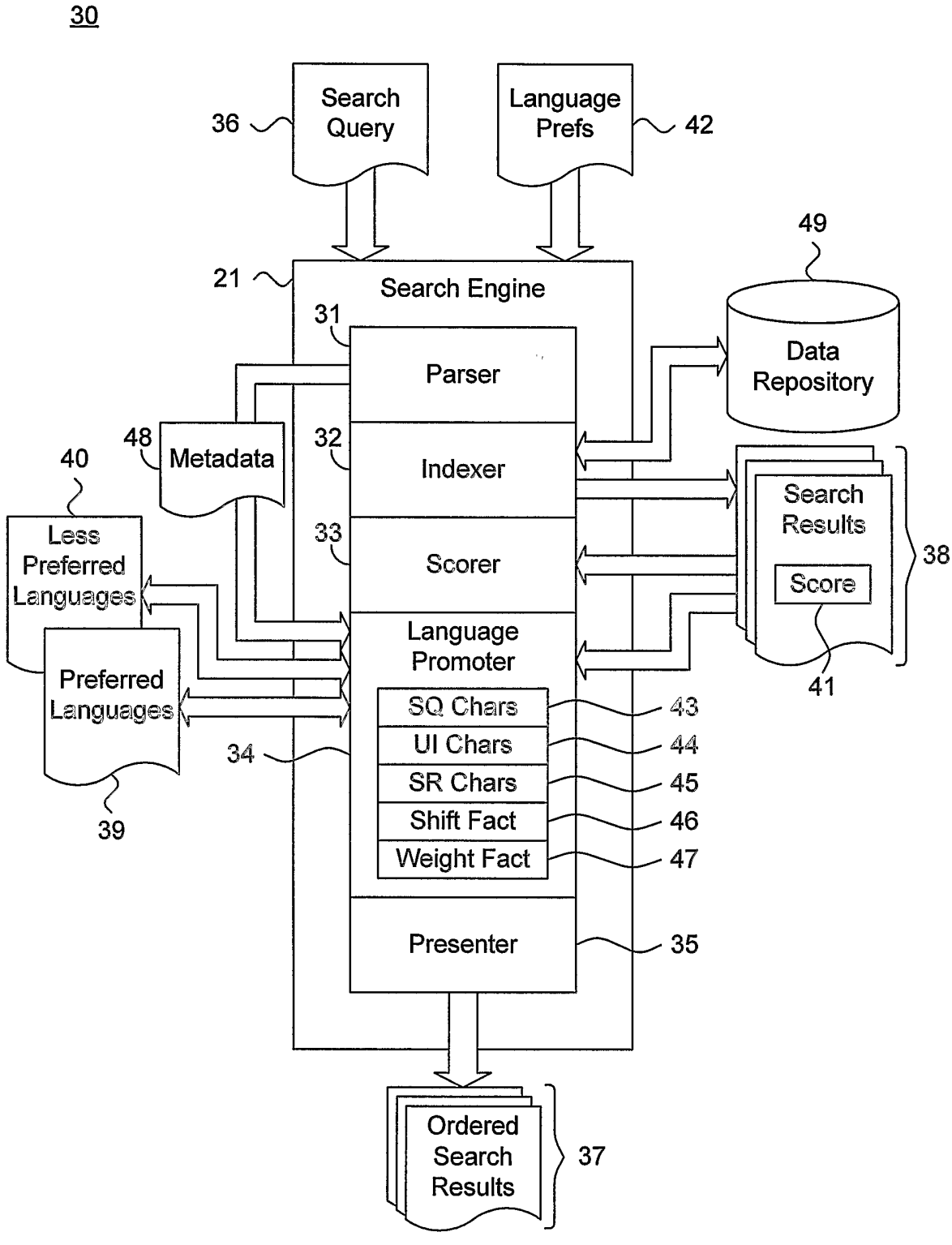

30
Search Query
36
Language Prefs
42
49
21
Search Engine
31
Parser
Data Repository
48
Metadata
32
Indexer
40
Search Results
Score
38
Less Preferred Languages
33
Scorer
Language Promoter
Preferred Languages
41
SQ Chars
43
UI Chars
44
34
SR Chars
45
Shift Fact
46
Weight Fact
47
39
Presenter
35
Ordered Search Results
37

Fig. 3.

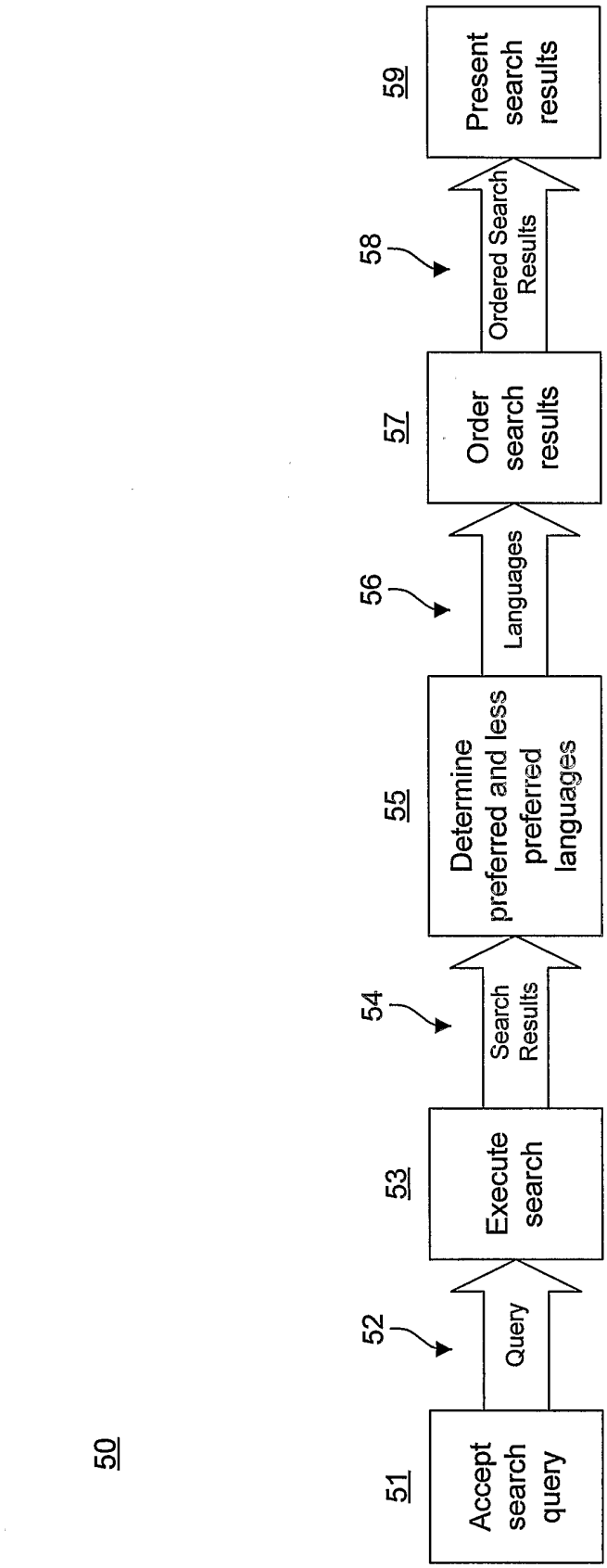

50
51
Accept search query
52
Query
53
Execute search
54
Search Results
55
Determine preferred and less preferred languages
56
Languages
57
Order search results
58
Ordered Search Results
59
Present search results

Fig. 4.

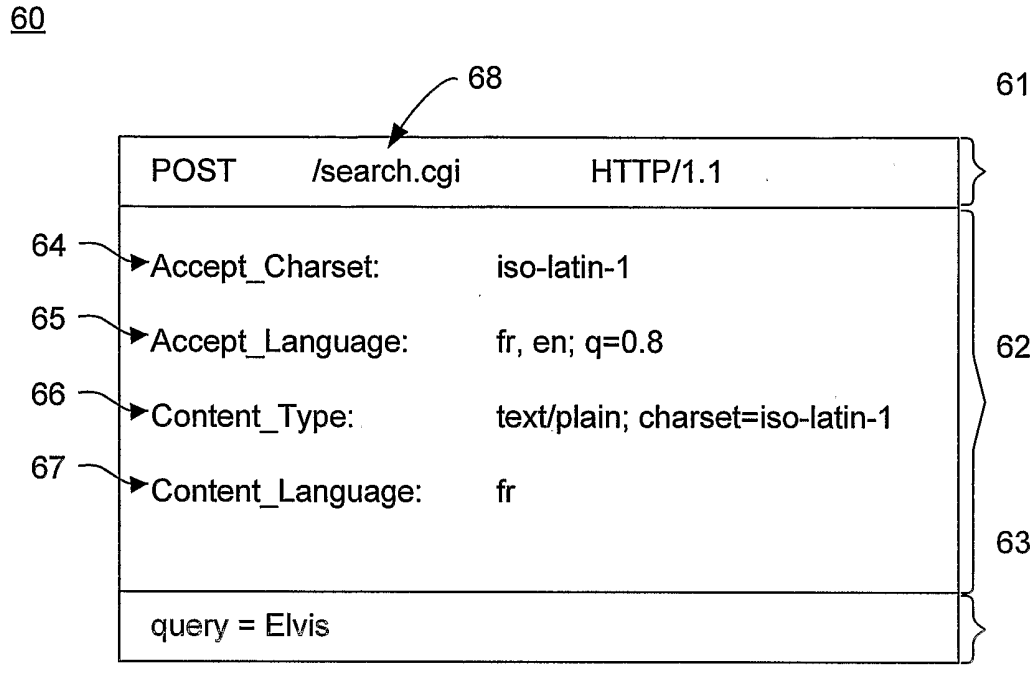

60
68
61
POST /search.cgi HTTP/1.1
64
Accept_Charset: iso-latin-1
65
Accept_Language: fr, en; q=0.8
66
Content_Type: text/plain; charset=iso-latin-1
67
Content_Language: fr
62
63
query = Elvis

Fig. 5.

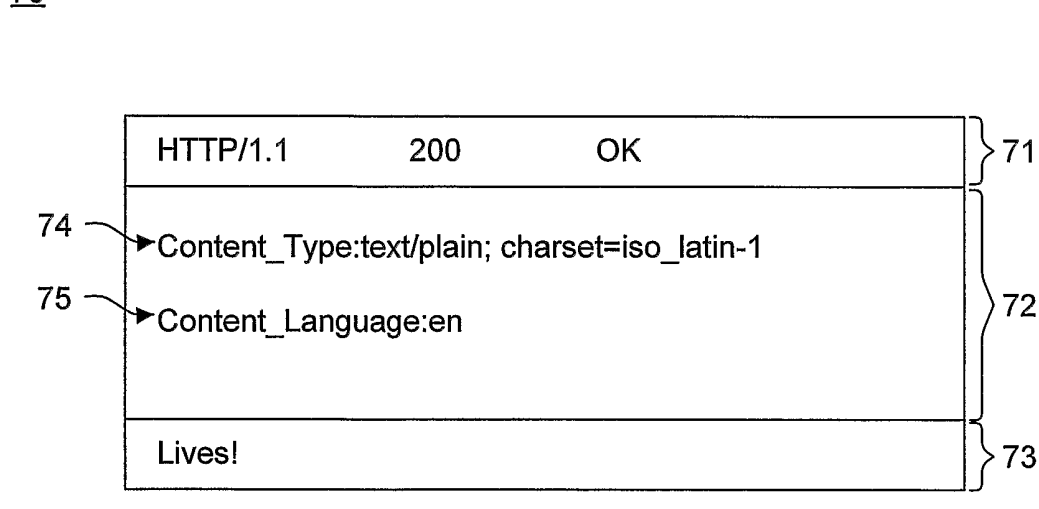

70
HTTP/1.1 200 OK
71
74
Content_Type:text/plain; charset=iso_latin-1
75
Content_Language:en
72
Lives!
73

Fig. 6.

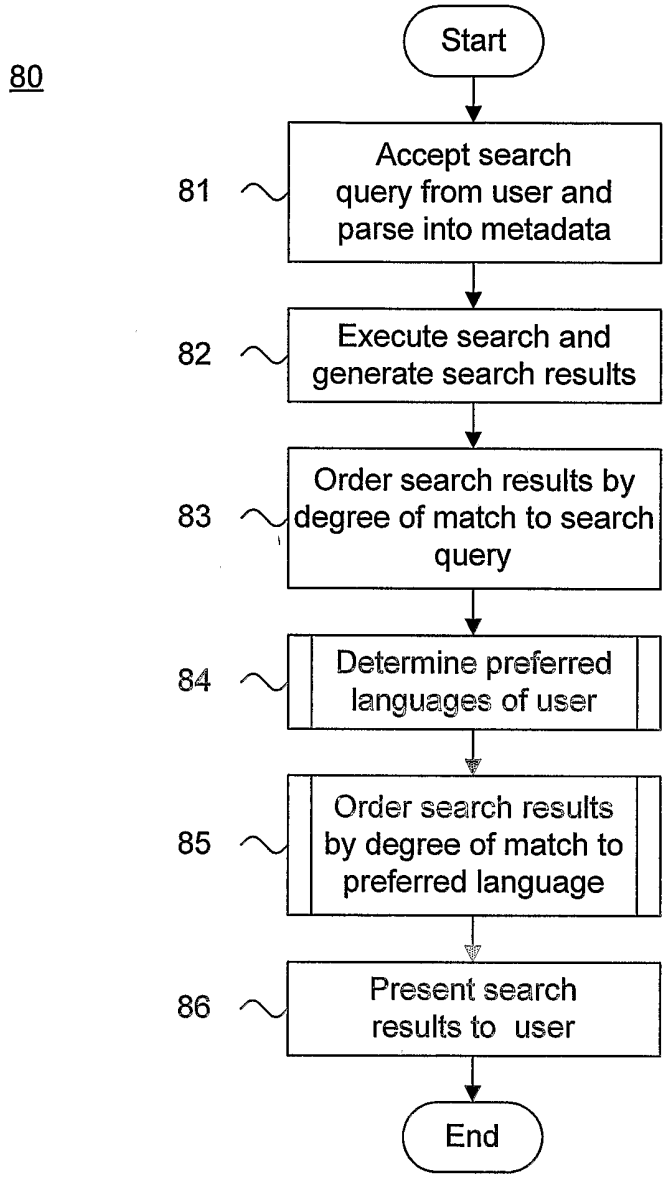

80
Start
81
Accept search query from user and parse into metadata
82
Execute search and generate search results
83
Order search results by degree of match to search query
84
Determine preferred languages of user
85
Order search results by degree of match to preferred language
86
Present search results to user
End

Fig. 7.

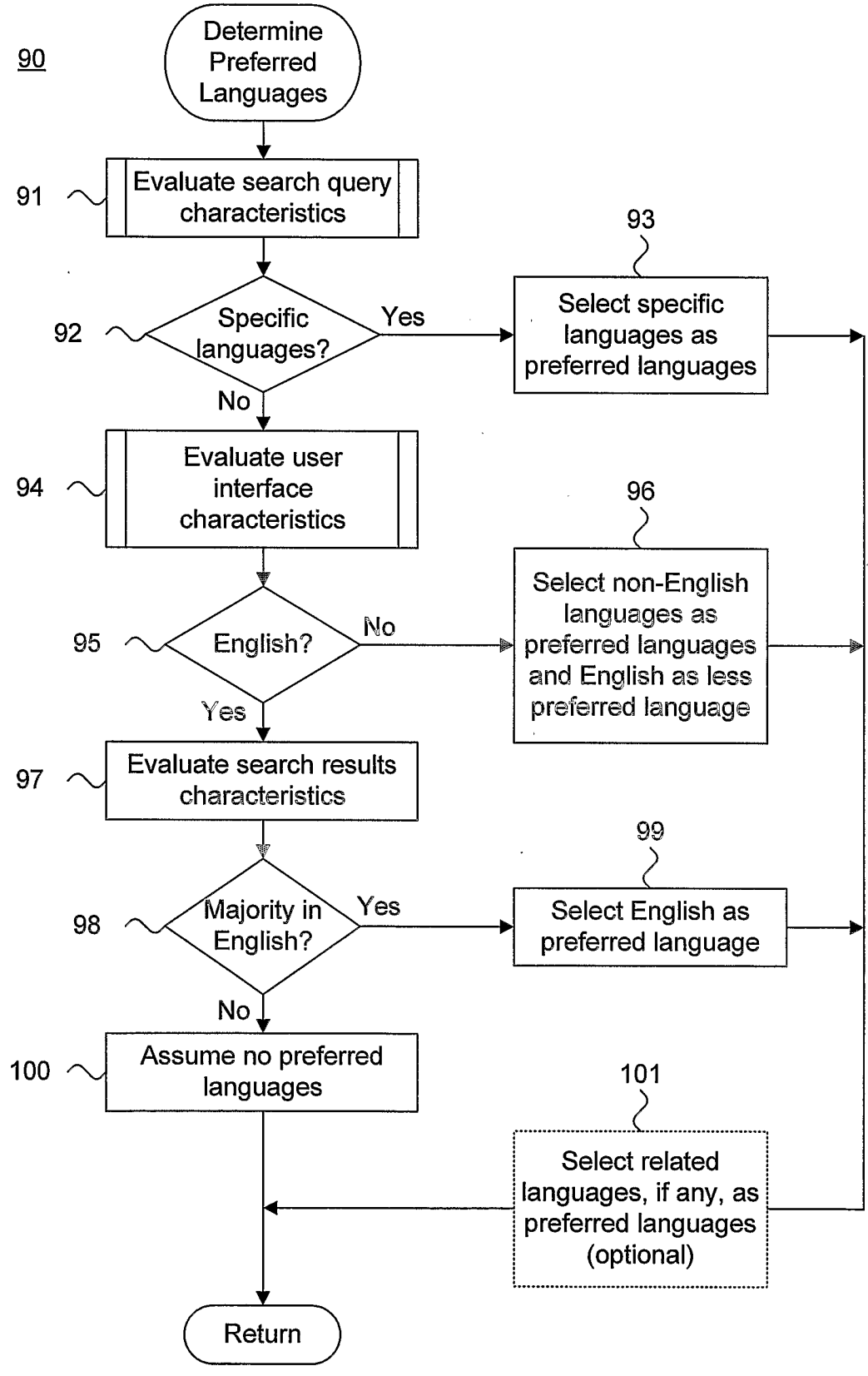

90
Determine Preferred Languages
91
Evaluate search query characteristics
92
Specific languages?
Yes
No
93
Select specific languages as preferred languages
94
Evaluate user interface characteristics
95
English?
No
Yes
96
Select non-English languages as preferred languages and English as less preferred language
97
Evaluate search results characteristics
98
Majority in English?
Yes
No
99
Select English as preferred language
100
Assume no preferred languages
101
Select related languages, if any, as preferred languages (optional)
Return

Fig. 8.

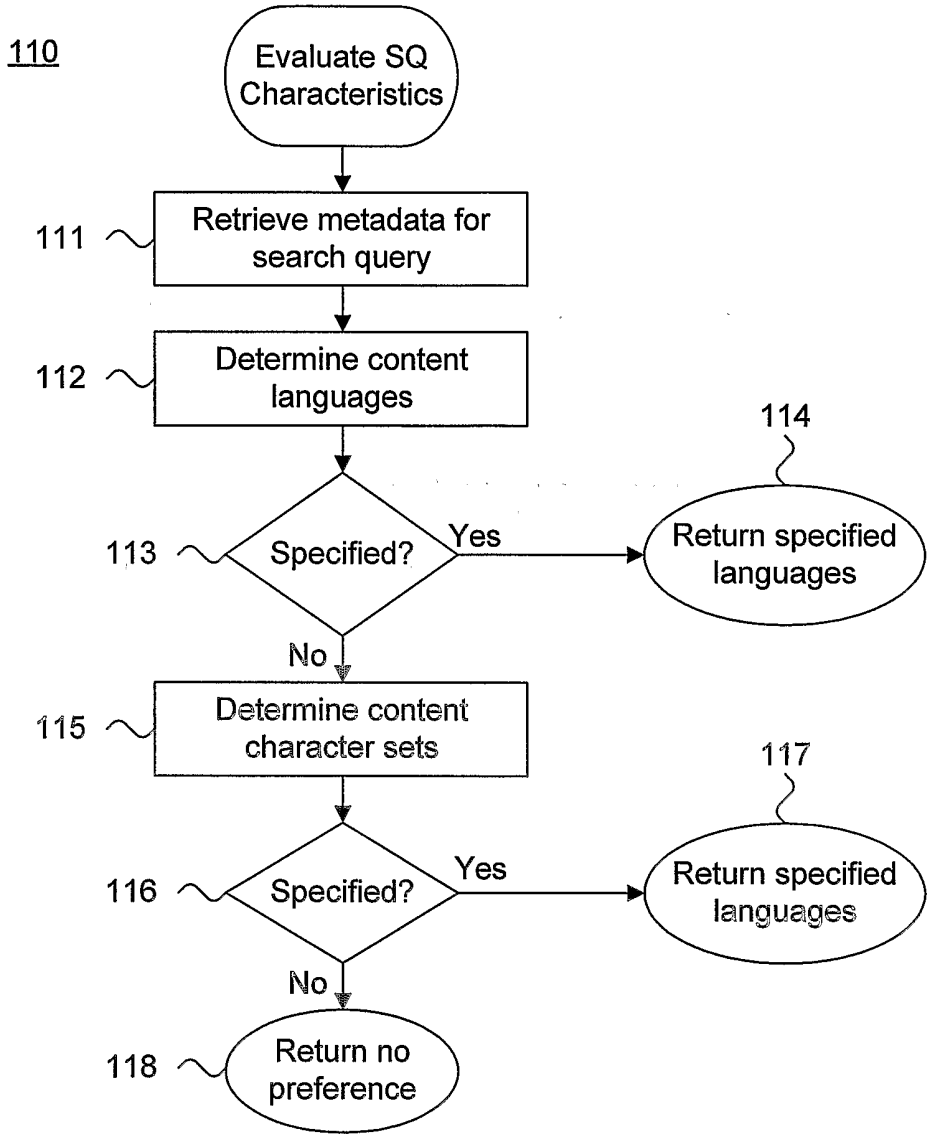

110
Evaluate SQ Characteristics
111
Retrieve metadata for search query
112
Determine content languages
113
Specified?
Yes
114
Return specified languages
No
115
Determine content character sets
116
Specified?
Yes
117
Return specified languages
No
118
Return no preference

Fig. 9.

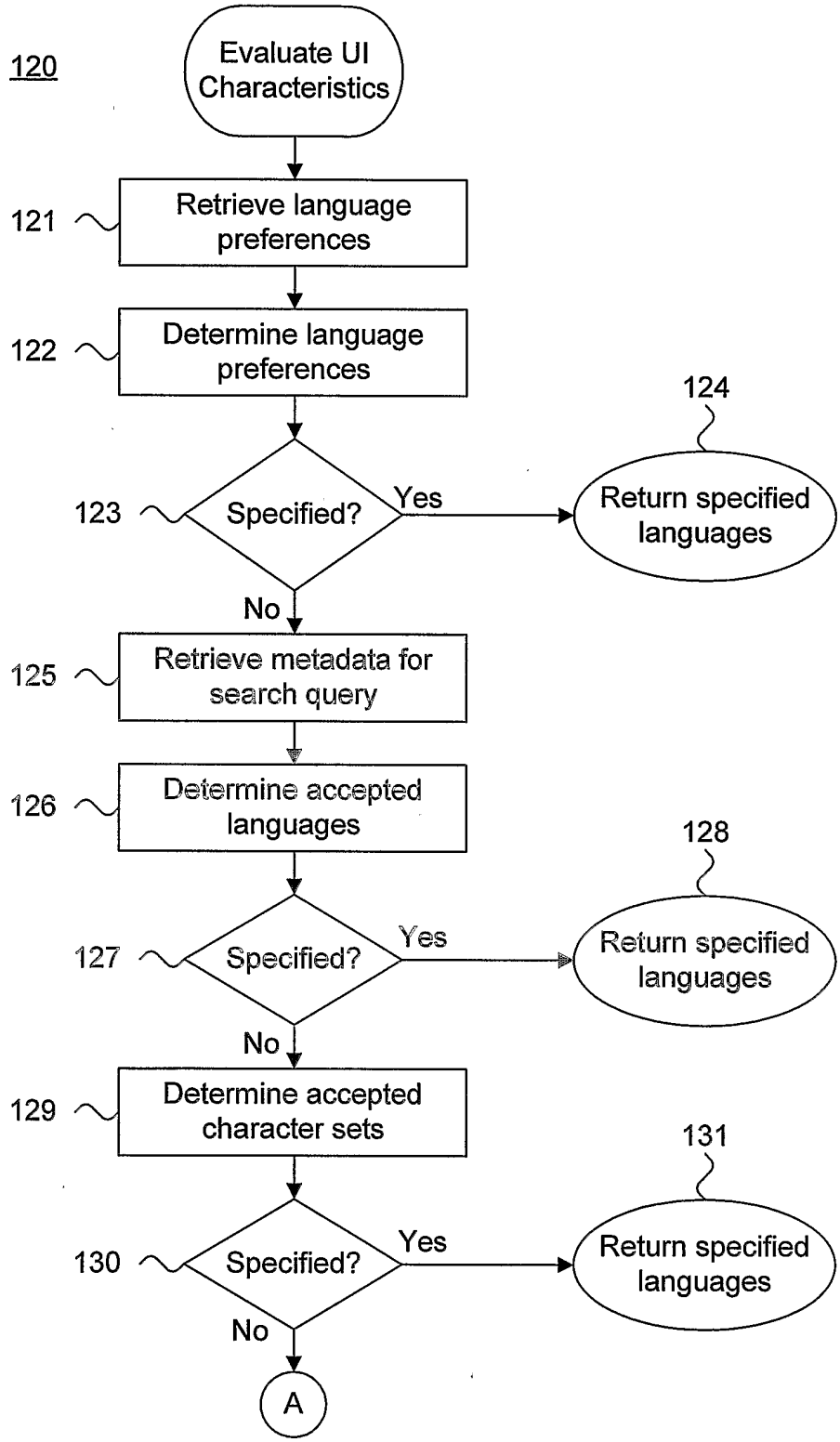

120
Evaluate UI Characteristics
121
Retrieve language preferences
122
Determine language preferences
123
Specified?
Yes
124
Return specified languages
No
125
Retrieve metadata for search query
126
Determine accepted languages
127
Specified?
Yes
128
Return specified languages
No
129
Determine accepted character sets
130
Specified?
Yes
131
Return specified languages
No
A

Fig. 9 (Cont.)

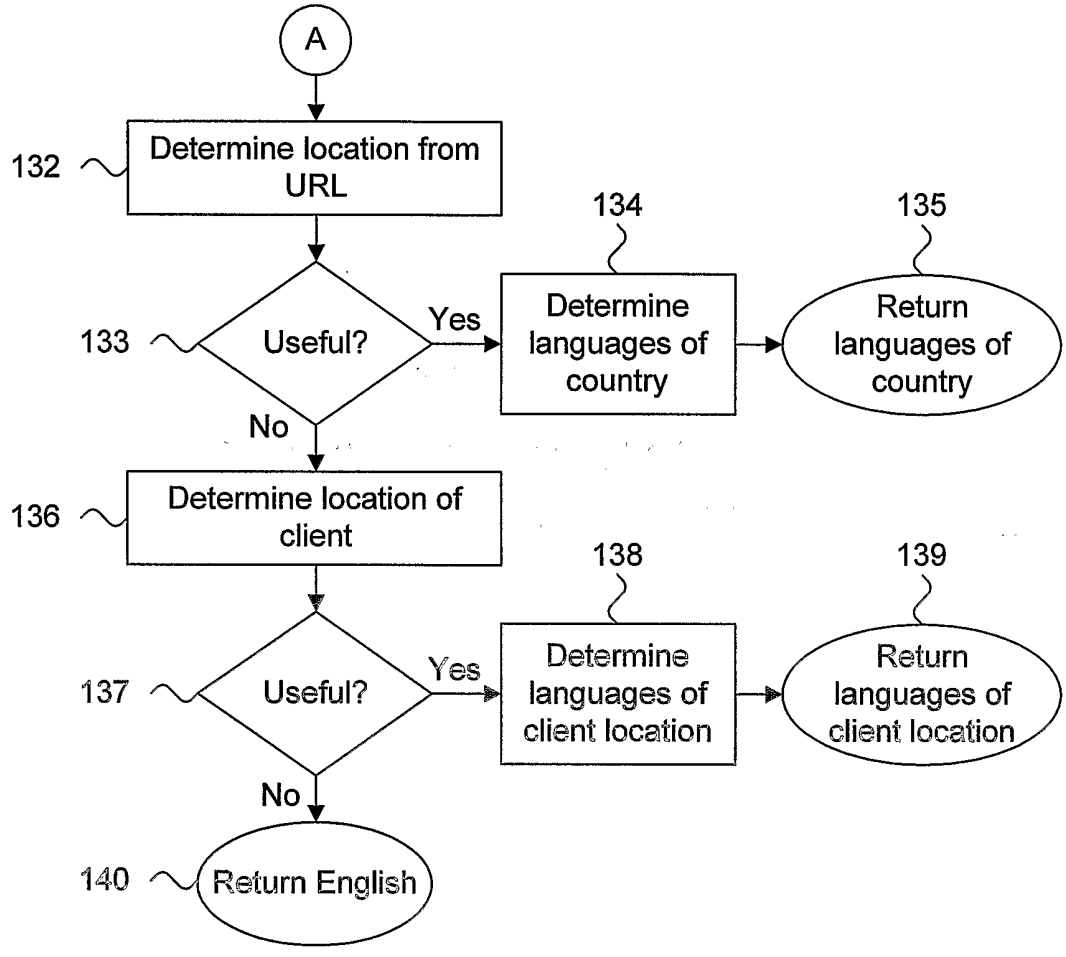

A
132
Determine location from URL
133
Useful?
Yes
134
Determine languages of country
135
Return languages of country
No
136
Determine location of client
137
Useful?
Yes
138
Determine languages of client location
139
Return languages of client location
No
140
Return English

Fig. 10.

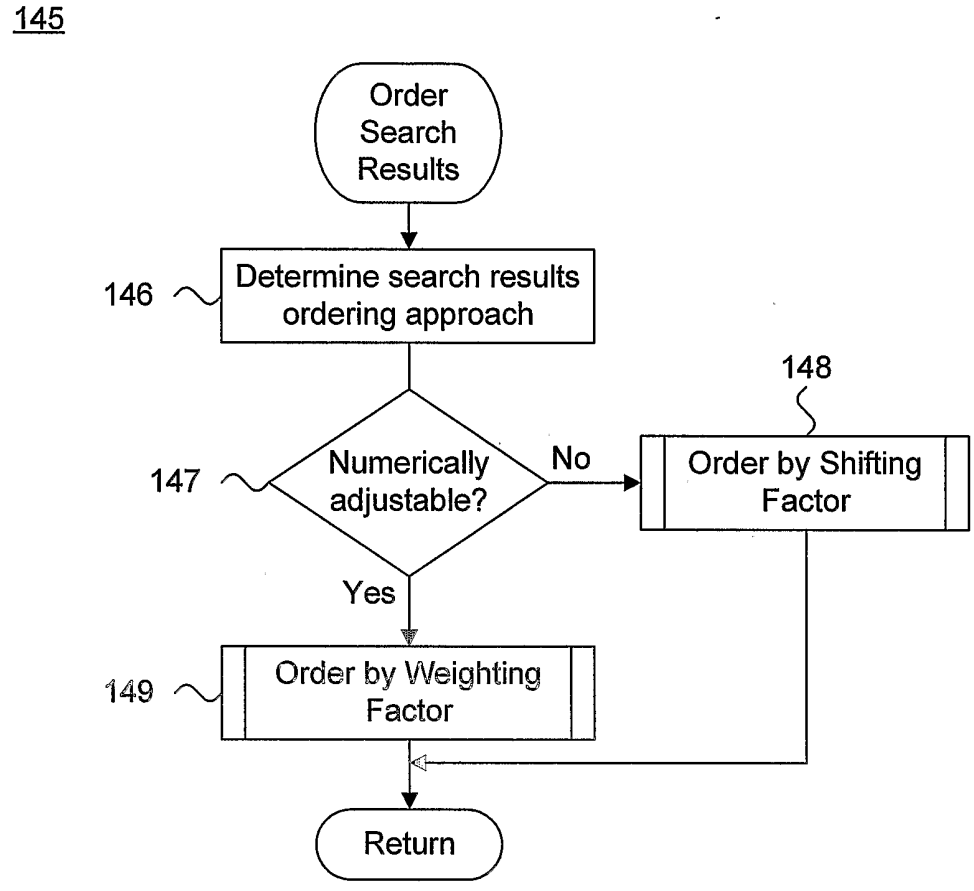

145
Order Search Results
146
Determine search results ordering approach
147
Numerically adjustable?
No
148
Order by Shifting Factor
Yes
149
Order by Weighting Factor
Return

Fig. 11.

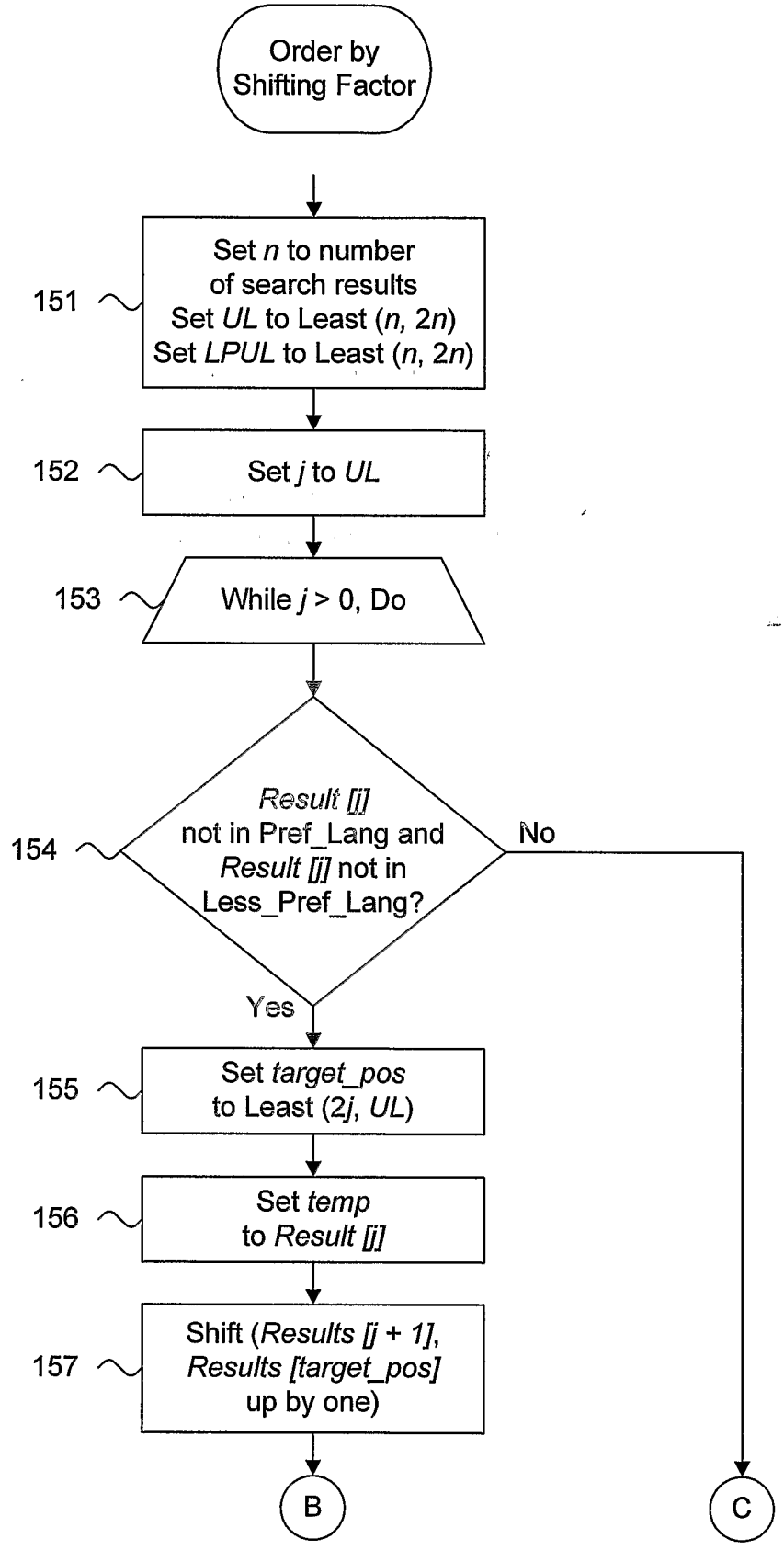

Order by Shifting Factor
150
151
Set n to number of search results
Set UL to Least (n, 2n)
Set LPUL to Least (n, 2n)
152
Set j to UL
153
While j > 0, Do
154
Result [j] not in Pref_Lang and Result [j] not in Less_Pref_Lang?
No
Yes
155
Set target_pos to Least (2j, UL)
156
Set temp to Result [j]
157
Shift (Results [j + 1], Results [target_pos] up by one)
B
C

Fig. 11 (Cont.)

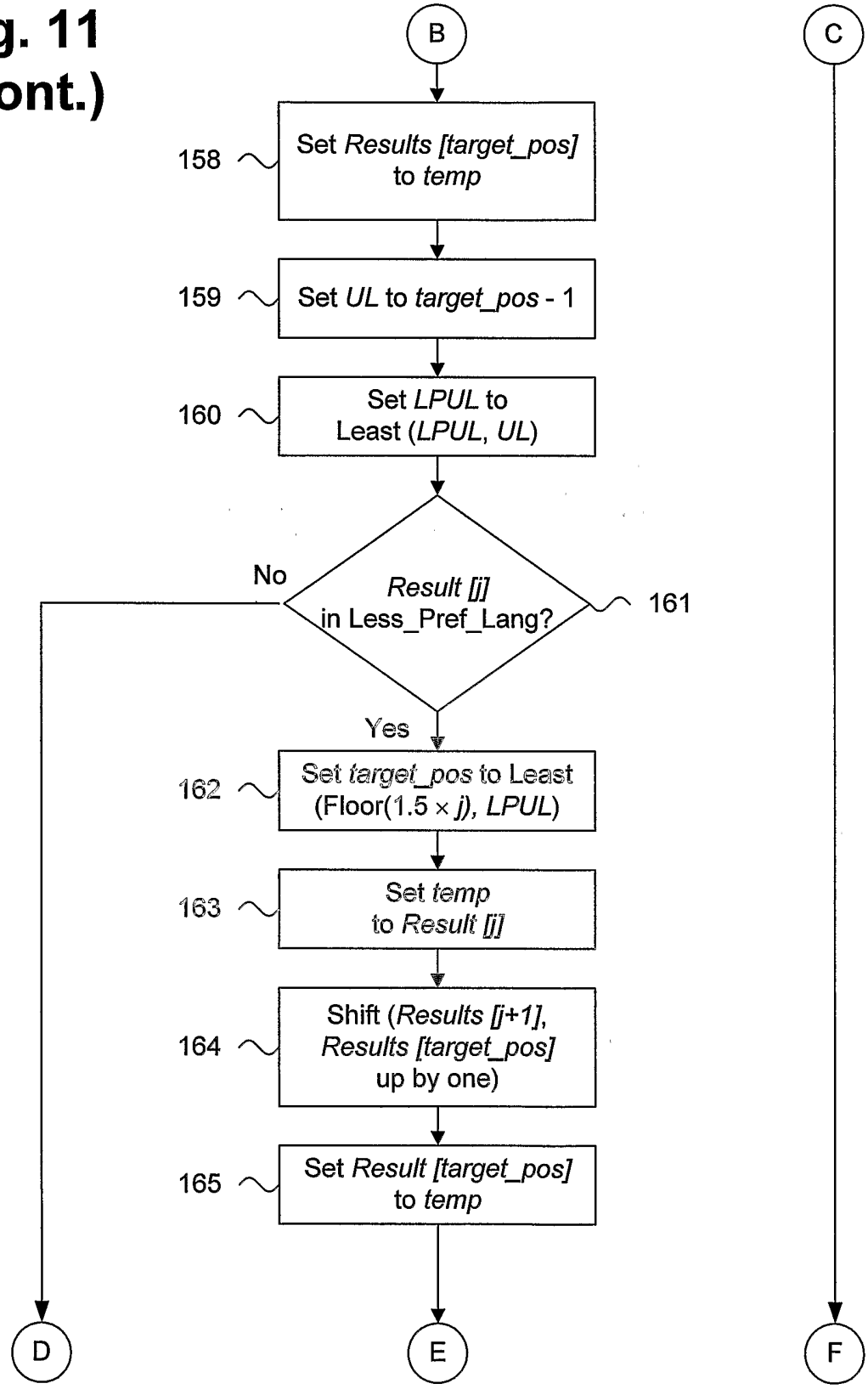

B
C
158
Set Results [target_pos] to temp
159
Set UL to target_pos - 1
160
Set LPUL to Least (LPUL, UL)
No
Result [j] in Less_Pref_Lang?
161
Yes
162
Set target_pos to Least (Floor(1.5 × j), LPUL)
163
Set temp to Result [j]
164
Shift (Results [j+1], Results [target_pos] up by one)
165
Set Result [target_pos] to temp
D
E
F

Fig.11 (Cont.)

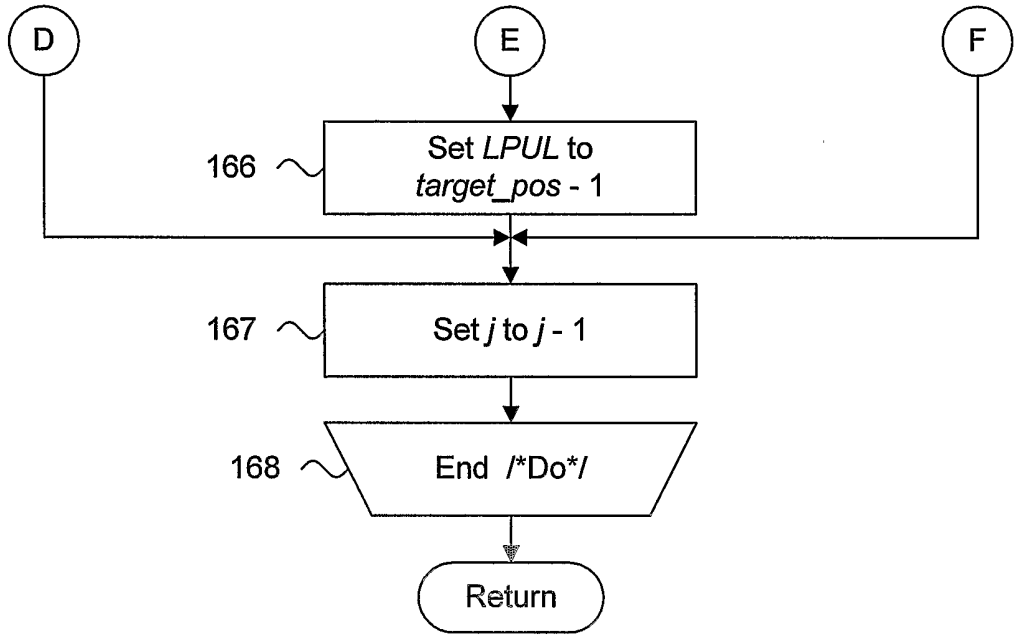
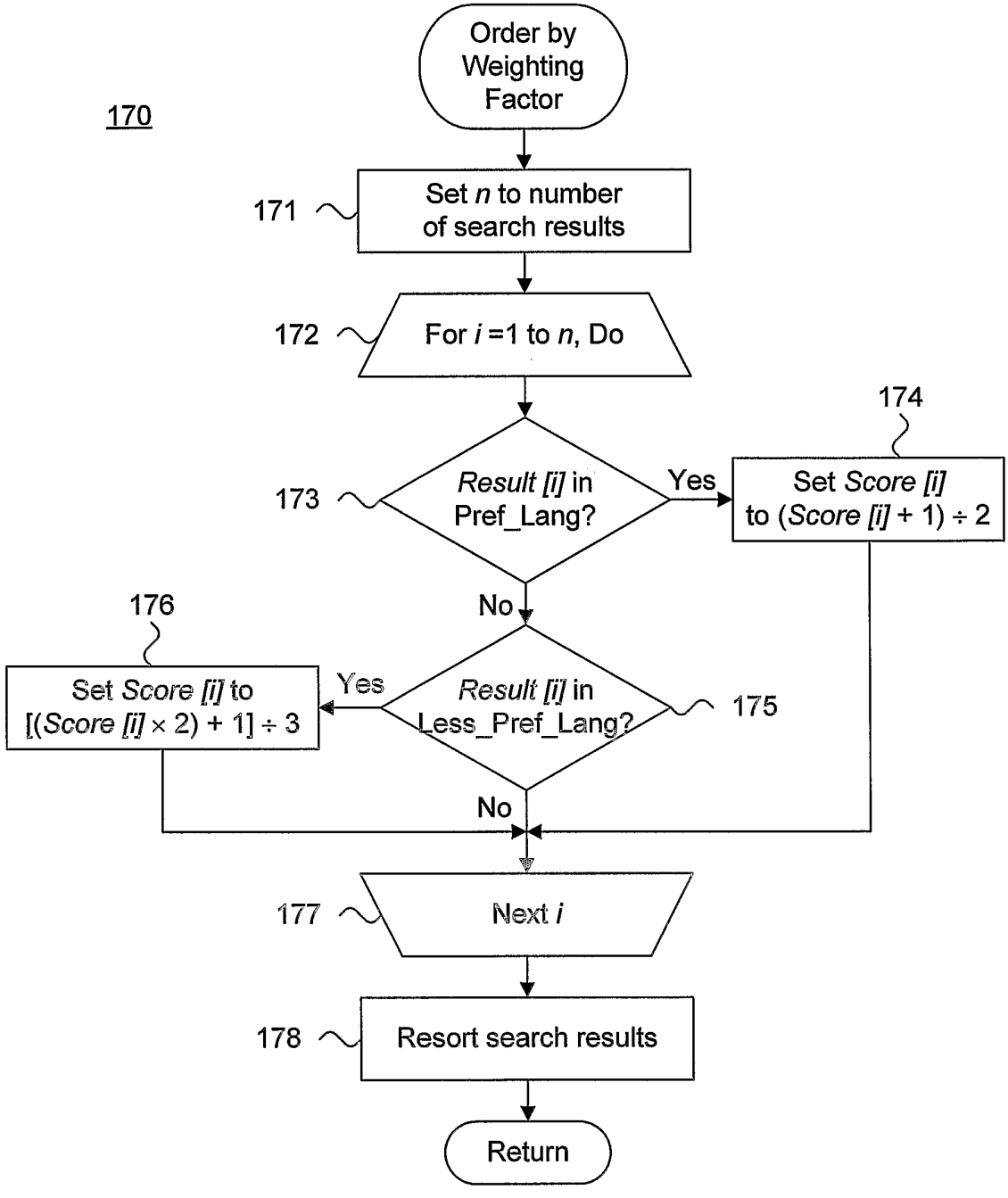

D
E
F
166
Set *LPUL* to *target_pos* - 1
167
Set *j* to *j* - 1
168
End /*Do*/
Return

Fig. 12.


170
Order by Weighting Factor
171
Set n to number of search results
172
For i =1 to n, Do
173
Result [i] in Pref_Lang?
Yes
174
Set Score [i] to (Score [i] + 1) ÷ 2
No
175
Result [i] in Less_Pref_Lang?
Yes
176
Set Score [i] to [(Score [i] × 2) + 1] ÷ 3
No
177
Next i
178
Resort search results
Return