

The opinion in support of the decision being entered today was *not* written for publication and is *not* binding precedent of the Board.

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

Ex parte DESIREE D. G. GOSBY and KEITH I. ITO

Appeal 2007-3941
Application 10/334,370
Technology Center 2100

Decided: September 28, 2007

Before ANITA PELLMAN GROSS, MAHSHID D. SAADAT,
and SCOTT R. BOALICK, *Administrative Patent Judges*.

SAADAT, *Administrative Patent Judge*.

STATEMENT OF THE CASE

This is a decision on appeal under 35 U.S.C. § 134(a) from the Examiner's Final Rejection of claims 1-8, 20, and 22-28, which are all of the claims pending in this application, as claims 9-19 and 21 have been canceled. We have jurisdiction under 35 U.S.C. § 6(b).

We affirm-in-part and enter a new ground of rejection pursuant to the provisions of 37 CFR § 41.50(b).

Appellants invented a method and system for document analysis and retrieval by using an algorithm for determining categories and document links relating to a given document (Specification 3). An understanding of the invention can be derived from a reading of independent claim 1 and dependent claim 5, which are reproduced as follows:

1. A method for document analysis and retrieval, comprising the steps of:
 - receiving a document having text therein from a host of a first computing system;
 - generating document keys associated with said text from analysis of said text, each said document key selected from the group consisting of a keyword of said text and a keyphrase of said text;
 - providing a document taxonomy having categories, each category having category keys, each said category key selected from the group consisting of a keyword of said category and a keyphrase of said category;
 - comparing the category keys of each category with said document keys to make a determination of a distance between the document and each category as a measure of how close the document is to each category; and
 - returning a subset of said categories to said host, wherein said subset of said categories reflects said determination.

5. The method of claim 1, wherein said comparing comprises computing said distance for each category as a dot product of a vector of said document keys and a vector of said category keys for each category.

The Examiner relies on the following prior art references:

Yanagihara	US 5,715,443	Feb. 3, 1998
Brown	US 2002/0099730 A1	Jul. 25, 2002

The rejections as presented by the Examiner are as follows:

1. Claims 1, 3-8, 20¹, and 23-28 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Brown.
2. Claims 2 and 22 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Brown and Yanagihara.

ISSUES

1. Under 35 U.S.C § 102(e), with respect to the appealed claims 1, 3-8, 20, and 23-28, does Brown anticipate the claimed subject matter by teaching all of the claimed limitations?
2. Under 35 U.S.C § 103(a), with respect to the appealed claims 2 and 22, would the ordinarily skilled artisan have found it obvious to modify Brown with Yanagihara to render the claimed invention unpatentable?

FINDINGS OF FACT

The following findings of fact (FF) are relevant to the issues involved in the appeal and are believed to be supported by a preponderance of the evidence.

1. Appellants' method claim 1 requires "receiving a document having text therein from a host of a first computing system" and "returning a subset of said categories to said host" after performing additional process

¹ Claim 21 is advertently listed in place of claim 20 in the statement of the rejection in both the Final Rejection and the Examiner's Answer.

steps. These process steps involve analyzing the document which may be performed outside a computer using no more than human intelligence.

2. Dependent claims 3 and 4 require that the steps of the method recited in their base claim be performed by the first and second computing devices, respectively.

3. Brown relates to an automatic text classification system which extracts words and word sequences from the texts to be analyzed and compares them with training data having a measure of probability with respect to a plurality of qualities. Each plurality of qualities may be represented by an axis, whose two end points correspond to mutually exclusive characteristics. Based on the comparison, the texts to be analyzed are then classified in terms of the plurality of qualities (Abstract).

4. Brown provides a system and method of generating classification data for text. The method comprises: identifying semantic content bearing lexical units in the data representing the text to be classified and determining classification data as a score for the text to be classified with respect to each of a plurality of qualities. Brown does so by comparing the identified lexical units with stored lexical units having a distribution of lexical scores associated therewith for each of a plurality of qualities (§ 0018 - § 0024).

5. Brown's training system comprises two parts - first, a classification of a plurality of pre-selected training texts in terms of each of a plurality of qualities and second, an automatic text analysis of each of the classified training texts. The object of the training system is to generate an output of singles, doubles, and triples of word stems and word stem

sequences together with a value on one or more axes to enable classification of subsequently-analyzed documents that contain the same words or combinations of words (§ 0051).

6. Thus, the automated classification process operates to determine scores for axes for documents based on extreme words and their synonyms and antonyms that are determined on an iterative basis. This avoids human subjective input that may give inaccurate retrieval results (§ 0064).

7. The result of the classification process is a series of scores (i.e., one on each axis) for each of the training texts. The output is illustrated schematically in Figure 5. Associated with each Training Text (illustrated by a dotted line) is a table or Score Table ST. The Score Table shown comprises two columns, namely an axis number and a score for each axis. Well known memory management techniques can be used to efficiently store the information. For example, a document number could simply be followed by n scores in a data array, thereby eliminating the storage of the axis identification numbers (§ 0065).

8. Brown generates a word stem and word stem sequences that are stored in association with the appropriate group. Using the example of the Happy-Sad axis, the stem “happi” will be expected to occur most frequently in group G0 of this axis. Thus, when this word stem “happi” is found in a new text the training data can be used to provide an indication that the document should be placed in one of the groups G0 on the Happy-Sad axis (§ 0093-0097).

9. The next step in the process is determining a score for each word stem and word stem sequence. This is carried out on a statistical basis. One example of a calculation of the likelihood or probability of occurrence of each of the stem words, doubles, and triples will now be described. It should be noted that, while a mathematical probability is given in the following examples, this need not be the case in practice (§ 0098).

10. The classification system processes texts in the same way as the training texts to identify word stems and their count, which are determined by a score (§ 0112).

11. For each axis, the probability of the new text belonging to each group on the axis is calculated (§ 0125). This relates the probability of the text being allocated to a particular group on each axis on the basis of the training data and the text being classified. This is performed by multiplying (for every word) the probabilities of that word occurring in a document that is allocated to that group (based on the training data) (§ 0126).

12. Having determined the differences using the split-merge-compare algorithm for the original training data, the classifications and word stem data for texts that were determined to give scores of high confidence are added to the original training data to provide modified training data, which is compared with the differences generated for the original training data (§ 0157).

13. Brown discloses different methods for comparison between scores. As depicted in Figure 13, the hierarchical structure of a classification tree is illustrated. In this embodiment the qualities or axes have extreme values indicating how much the document is concerned with a

topic. Thus the extremes can be simply YES and NO. This hierarchical structure requires 4 classifiers having 4 different sets of training data, which provide 3 or 2 qualities or axes for which the documents are given scores by automatic or manual classification. The word stems and word stem sequences in the subset of documents are identified to obtain the training data which will give scores for the axes and the associated distribution of word stem and word stem sequence scores across the groups as illustrated in Figure 7. Thus the highest score for one of the qualities or axes will determine the classification assigned (§ 0162).

14. Figure 15 of Brown shows the classification of texts and matching them to the user requirements are carried out remotely at the premises of a cable television distributor (§ 0168).

15. Yanagihara relates to document searching wherein stop words used in the search queries are not considered in data searching and document retrieving processes (col. 14, ll. 19-49).

PRINCIPLES OF LAW

1. Scope of claims

Absent an express intent to impart a novel meaning to a claim term, the words take on the ordinary and customary meanings attributed to them by those of ordinary skill in the art. *Brookhill-Wilk 1, LLC v. Intuitive Surgical, Inc.*, 334 F.3d 1294, 1298, 67 USPQ2d 1132, 1135-36 (Fed. Cir. 2003). The claim construction analysis begins with the words of the claim. *See Vitronics Corp. v. Conceptronic, Inc.*, 90 F.3d 1576, 1582, 39 USPQ2d, 1573, 1576 (Fed. Cir. 1996). Claims will be given their broadest reasonable

interpretation consistent with the specification, and limitations appearing in the specification are not to be read into the claims. *In re Etter*, 756 F.2d 852, 858, 225 USPQ 1, 5 (Fed. Cir. 1985).

2. *Anticipation*

A rejection for anticipation under section 102 requires that each and every limitation of the claimed invention be disclosed in a single prior art reference. *See In re Paulsen*, 30 F.3d 1475, 1478-79, 31 USPQ2d 1671, 1673 (Fed. Cir. 1994). Anticipation of a claim requires a finding that the claim at issue reads on a prior art reference. *Atlas Powder Co. v. IRECO Inc.*, 190 F.3d 1342, 1346, 51 USPQ2d 1943, 1945 (Fed. Cir. 1999) (quoting *Titanium Metals Corp. v. Banner*, 778 F.2d 775, 781, 227 USPQ 773, 778 (Fed. Cir. 1985)).

3. *Obviousness*

The test for obviousness is what the combined teachings of the references would have suggested to one of ordinary skill in the art. *See In re Kahn*, 441 F.3d 977, 987-988, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006), *In re Young*, 927 F.2d 588, 591, 18 USPQ2d 1089, 1091 (Fed. Cir. 1991) and *In re Keller*, 642 F.2d 413, 425, 208 USPQ 871, 881 (CCPA 1981).

The Examiner can satisfy this burden by showing some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness. *KSR Int'l. v. Teleflex Inc.*, 127 S. Ct. 1727, 1741, 82 USPQ2d 1385, 1396 (2007) (*citing In re Kahn*, 441 F.3d 977, 988, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006)).

ANALYSIS

1. 35 U.S.C. § 102 Rejection

Claims 1 and 20

Appellants agree with the Examiner's findings in Brown with respect to the steps of "receiving a document having text therein" and "generating document keys associated with said text" such that the text in both steps are taught to be the same text (Reply Br. 6).

With respect to the claim term "providing a document taxonomy," Appellants assert that a "classified document" disclosed in Brown is not the same as a "document taxonomy," (Br. 6-7; Reply Br. 6-9). The Examiner relies on ¶¶ 0092-97 of Brown and argues that the method of providing categories for the word stems and word stem sequences and storing the procedure in association with each group is the same as the document taxonomy (Answer 7-8). Appellants point to page 6 of their Specification for a description of taxonomy. Actually, consistent with Appellants' own disclosure, we find that Brown considers "films" as the document taxonomy, whereas categories form axes such as "Happy-Sad" and word stems get scores for each group within the axis (FF 8). We also note that Appellants' claim 1 recites the term "document taxonomy" only in the context of defining categories and category keys associated therewith. We also observe that the claims as a whole, do not recite any limitation that attributes any functionality to the document taxonomy.

Regarding the feature of "comparing the category keys ... to make a determination of a distance between document keys," Appellants argue that

Brown merely compares lexical units which are words and phrases and not category keys (Reply Br. 9-10). The Examiner refers to Brown's series of scores and argues that they are results of comparison (Answer 9-10). We agree with the Examiner's position that Brown's determination of scores for each category results in a determination of a distance or difference between the training data and the text to be classified (FF 9). Taking the broadest reasonable interpretation of the claim, we find that any comparison that provides a difference of scores would read on the claimed term "distance," which indicates the likelihood or frequency of occurrence of the word stem in the data (FF 8-10).

Appellants further argue that Brown does not disclose the claimed requirement that the documents are received from a host of a first computing system while the subset of the categories is returned to the same host (Reply Br. 11). The Examiner responds by pointing to the portion of Brown describing the hierarchical structure of the classification tree and argues that by classifying the document the subset of the categories is returned to the host (Answer 11). We agree with the Examiner's argument since the classification of a text is achieved by returning the scores and the associated distribution of word stem and word stem sequence scores across the groups (FF 12-13). As such Brown determines the classification for the text according to the highest score associated with each word stem.

Appellants do not present any separate arguments in support of the patentability of claims 3, 20, and 23 and allow these claims to fall with claim 1 (Br. 9). Therefore, in view of the analysis above and to the extent claimed, we find that Brown prima facie anticipates claims 1 and 20, as well as claims 3 and 23, dependent thereon.

Claims 4 and 24

Appellants contend that the portions of Brown relied on by the Examiner require the host of the first computer to be at the “user GUI” in the second computer system and thus, makes it impossible for the user to be both the first and the second computer (Reply Br. 12-13). The Examiner points to Figure 15 of Brown and argues that the main host is the first computing system which provides the text to the user, which, as the second computer system, performs the claimed method steps (Answer 11). We agree with the Examiner and find the Examiner’s position reading the claimed second computing system on the user in Brown to be reasonable (FF 14).

Claims 5-8 and 25-28

With respect to claims 5 and 25, Appellants further contend that the claimed “dot product of a vector” cannot be read on Brown's “multiplying” of the probabilities of the word occurring in the document (Reply Br. 13). The Examiner’s arguments (Answer 11-12) do not explain how the disclosed multiplying operation of probabilities may be same as the dot product of a vector, as no vector has been identified in Brown. We observe that claims 6

and 26 also recite a “dot product of a vector of document weights and a vector of said category weights,” which is not taught by Brown. Therefore, we agree with Appellants and find that, since all the claimed limitations are not taught, Brown cannot anticipate claims 5, 6, 25, and 26, as well as claims 7, 8, 27, and 28, dependent thereon.²

2. *35 U.S.C. § 103 Rejection*

With respect to the rejection of claims 2 and 22, Appellants argue that Yanagihara may not be combined with Brown since no suggestion was presented for the combination and the reference has no teaching related to displaying text in a non-text format nor generating tokens of the text to be used in generating the document keys (Reply Br. 16-19). The Examiner appears to read the claimed “extracting text from said document” as copying the text from the document (Answer 12). The Examiner further relies on the search strategy of Yanagihara depicted in Figure 4b for teaching generating tokens of said text (*id.*).

Again, we agree with the Examiner’s position since the claims are broad enough to encompass copying the text as the claimed extracting text. Additionally, the claimed “non-text format” reads on any graphic or other formats that are not text. We also find that since using tokens is known in the art, one of ordinary skill in the art would have found it obvious to use tokens of Yanagihara in Brown for removing stop words from the search (FF 15) and performing the queries more efficiently by using tokens.

² It appears that words are missing from lines 4 and 8 of claims 6 and 26, after the word “importance.”

CONCLUSION

On the record before us, Appellants have failed to show that the Examiner has erred in rejecting claims 1-4, 20, and 22-24. However, we found error in the Examiner's rejection of claims 5-8 and 25-28. In view of our analysis above, we sustain the 35 U.S.C. § 102 rejection of claims 1, 3, 4, 20, 23, and 24 over Brown and the 35 U.S.C. § 103 rejection of claims 2 and 22 over Brown and Yanagihara. However, we do not sustain the 35 U.S.C. § 102 rejection of claims 5-8 and 25-28.

NEW GROUND OF REJECTION

We enter the following new rejection of claims 1, 2, and 5-8 under the provisions of 37 CFR § 41.50 (b).

Claims 1, 2, and 5-8 are rejected under 35 U.S.C. § 101 as being directed to non-statutory subject matter.

In performing the method steps of claim 1, there is no requirement that a computer be used. The only recitation of "computer" or "host" relates to the source or destination of data transmission. For example, a document is received from a host and a subset of said categories is returned to the host. The claim is merely drawn to "disembodied abstract ideas," which do not have any "real world effect" until they are implemented. The absence of any transformation of physical subject matter according to the definition of a process under 35 U.S.C. § 101, places this claim on the other side of the line defining statutory subject matter. A case involving this issue is presently on appeal to the Federal Circuit: *In re Bilski*, No. 2007-1130 (to be argued Oct. 1, 2007). Additionally, in performing the method steps of claim 1, there is

no requirement that a computer be used. Therefore, the claimed subject matter may be performed using only human intelligence, which has recently been held to be non-statutory. *In re Comiskey*, No. 2006-1286, Slip Op. at 21 (Fed. Cir. Sep. 20, 2007).

For the above reasons, we find that claim 1, as well as claims 2 and 5-8, which depend thereon, recite non-statutory subject matter. Accordingly, claims 1, 2, and 5-8 are rejected under 35 U.S.C. § 101.

DECISION

The decision of the Examiner rejecting claims 1, 3-8, 20, and 23-28 under 35 U.S.C. § 102 based on Brown is reversed with respect to claims 5-8 and 25-28 and affirmed with respect to claims 1, 3, 4, 20, 23, and 24. The 35 U.S.C. § 103 rejection of claims 2 and 22 based on Brown and Yanagihara is affirmed.

This decision contains a new ground of rejection pursuant to 37 CFR § 41.50(b) (effective September 13, 2004, 69 Fed. Reg. 49960 (August 12, 2004), 1286 Off. Gaz. Pat. Office 21 (September 7, 2004)). 37 CFR § 41.50(b) provides "[a] new ground of rejection pursuant to this paragraph shall not be considered final for judicial review."

37 CFR § 41.50(b) also provides that the appellant, WITHIN TWO MONTHS FROM THE DATE OF THE DECISION, must exercise one of the following two options with respect to the new ground of rejection to avoid termination of the appeal as to the rejected claims:

- (1) *Reopen prosecution*. Submit an appropriate amendment of the claims so rejected or new evidence relating to the claims so rejected, or both, and have the matter reconsidered by the examiner, in which event the proceeding will be remanded to the examiner. . . .

Appeal 2009-3941
Application 10/334,370

(2) *Request rehearing.* Request that the proceeding be reheard under § 41.52 by the Board upon the same record. . . .

No time period for taking any subsequent action in connection with this appeal may be extended under 37 C.F.R. § 1.136(a)(1)(iv).

AFFIRMED-IN-PART - 37 C.F.R. § 41.50(b)

pgc

SCHMEISER, OLSEN & WATTS
22 CENTURY HILL DRIVE
SUITE 302
LATHAM NY 12110