Internal distribution code:
(A) [ ] Publication in OJ
(B) [ - ] To Chairmen and Members
(C) [ - ] To Chairmen
(D) [ X ] No distribution

Datasheet for the decision
of 14 July 2017

Case Number: T 2418/12 - 3.5.07
Application Number: 05102959.3
Publication Number: 1587011
IPC: G06F17/30
Language of the proceedings: EN

Title of invention:
Related term suggestion for multi-sense queries

Applicant:
Microsoft Technology Licensing, LLC

Headword:
Related-term suggestion/MICROSOFT TECHNOLOGY LICENSING

Relevant legal provisions:
EPC Art. 56

Keyword:
Inventive step - all requests (no)

Decisions cited:
T 0154/04, T 1784/06, T 1928/06, T 1358/09, T 0042/10,
T 2230/10, T 1370/11, T 1734/11, T 2439/11, T 1742/12
DECISION of Technical Board of Appeal 3.5.07 of 14 July 2017

Appellant: Microsoft Technology Licensing, LLC
(Applicant)
One Microsoft Way
Redmond, WA 98052 (US)

Representative: Grünecker Patent- und Rechtsanwälte
PartG mbB
Leopoldstraße 4
80802 München (DE)

Decision under appeal: Decision of the Examining Division of the European Patent Office posted on 18 July 2012 refusing European patent application No. 05102959.3 pursuant to Article 97(2) EPC.

Composition of the Board:
Chairman: R. Moufang
Members: R. de Man
M. Jaedicke
Summary of Facts and Submissions

I. The applicant (appellant), which at the time was Microsoft Corporation, appealed against the decision of the Examining Division refusing European patent application No. 05102959.3.

II. With effect from 2 February 2015 the EPO registered a transfer of the application to Microsoft Technology Licensing, LLC.

III. The Examining Division decided that the subject-matter of claim 1 of the main request and of the then first, second and third auxiliary requests lacked inventive step in view of the following document:


IV. In the statement of grounds of appeal, the appellant maintained the main request, withdrew the first auxiliary request and maintained the second and third auxiliary requests considered in the decision under appeal as first and second auxiliary requests.

V. In a communication accompanying a summons to oral proceedings, the Board raised a number of clarity objections and expressed the preliminary view that the subject-matter of claim 1 of each request lacked inventive step both over a notorious general-purpose computer and over document D5.

VI. In a letter dated 14 June 2017, the appellant maintained its main request and first and second
auxiliary requests as main request and auxiliary requests II and III, respectively, and submitted new auxiliary requests I and IV.

VII. Oral proceedings were held on 14 July 2017. At the end of the oral proceedings, the chairman pronounced the Board's decision.

VIII. The appellant's final requests were that the decision under appeal be set aside and that a patent be granted on the basis of the claims of the main request filed with letter of 22 May 2012 or, alternatively, on the basis of the claims of one of auxiliary request I filed with letter of 14 June 2017, auxiliary request II filed with letter of 22 May 2012 as second auxiliary request, auxiliary request III filed as third auxiliary request during oral proceedings before the Examining Division on 28 June 2012 and auxiliary request IV filed with letter of 14 June 2017.

IX. Claim 1 of the main request reads as follows:

"A computer-implemented method for related term suggestion

characterized by

using a configurable threshold value to assign a historical query term (108) to have a high frequency of occurrence, so-called FOO, when this historical query term (108) occurs at least a threshold number of times, and to assign a historical query term (108) to have a low FOO, when this historical query term (108) occurs less than a threshold number of times;
generating term clusters (138) as a function of calculated similarity of term vectors (136), each term vector (136) being generated from search results associated with the term (108) which is a term of a set of high FOO historical query terms (116) previously submitted to a search engine (126), wherein the calculated similarity, \( \text{sim}(q_j, q_k) \), is determined as follows:

\[
\text{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{ji} \cdot w_{ki},
\]

wherein weight \( w \) for the \( i^{th} \) vector's \( j^{th} \) term is calculated as follows:

\[
w_{ji} = TF_{ij} \cdot \log(N / DF_j),
\]

wherein \( TF_{ij} \) represents term frequency, \( N \) is a total number of query terms, and \( DF_j \) is a number of extracted feature records that contain term \( j \);

responsive to receiving a term from an entity, evaluating the term in view of matches of the term (108) with terms in the term clusters (138) to identify one or more related term suggestions, wherein a multi-sense query comprises the term (108); and

returning at least one suggested term list (110), wherein the terms within each list are ordered by a combination of FOO and confidence value, wherein multiple suggested term lists are generated when the term matches terms in more than one term cluster, the lists are ordered by term cluster sizes, and wherein the confidence value is defined as:

\[
\text{dist}(q_j, q_k) = 1 - \text{sim}(q_j, q_k).
\]
X. Claim 1 of auxiliary request I reads as follows:

"A computer-implemented method for related term suggestion for a multi-sense query

characterized by

using a configurable threshold value to assign a historical query term (108) to have a high frequency of occurrence, so-called FOO, when this historical query term (108) occurs in historical queries at least a threshold number of times, and to assign a historical query term (108) to have a low FOO, when this historical query term (108) occurs in the historical queries less than a threshold number of times, wherein the historical queries include query terms previously submitted to a search engine (132);

generating term clusters (138) as a function of calculated similarity of term vectors (136), each term vector (136) being generated from search results associated with the term (108) which is a term of a set of high FOO historical query terms (116) previously submitted to a search engine (126), wherein the calculated similarity, \( \text{sim}(q_j, q_k) \), is determined as follows:

\[
\text{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{i,j} \cdot w_{i,k} ;
\]

wherein weight \( w \) for the \( i^{th} \) vector's \( j^{th} \) term is calculated as follows:

\[
w_{i,j} = TF_{i,j} \times \log(N / DF_j),
\]

wherein \( TF_{i,j} \) represents term frequency, i.e., a number of occurrences of term \( j \) in the \( i^{th} \) record, \( N \) is a
total number of query terms, and \( DF_j \) is a number of
extracted feature records that contain term \( j \);

responsive to receiving a term from an entity,
evaluating the term in view of matches of the term
(108) with terms in the term clusters (138) to identify
one or more related term suggestions, wherein a multi-
sense query comprises the term (108), wherein a match
is one of an exact match or a match with a small number
of variations such as singular/plural forms,
misspellings, punctuation marks; and

returning at least one suggested term list (110),
wherein the terms within each list are ordered by a
combination of FOO and confidence value, wherein
multiple suggested term lists are generated when the
term matches terms in more than one term cluster, the
lists are ordered by term cluster sizes, and wherein
the confidence value is defined as:

\[
dist(q_j,q_k) = 1 - \text{sim}(q_j,q_k).
\]

XI. Claim 1 of auxiliary request II differs from claim 1 of
the main request in that the following text has been
added at the end of the claim:

"wherein the term clusters are a first set of term
clusters (138), and wherein the method further
comprises:

determining that there is no match between the term and
the terms; and

responsive to the determining:
making a second set of term clusters from calculated similarity of term vectors (136), each term vector (136) being generated from search results associated with a set of low FOO historical query terms (122) previously submitted to the search engine (126); and

evaluating the term (108) in view of terms of the second set of term clusters (138) to identify one or more related term suggestions;

wherein making further comprises:

identifying the low FOO historical query terms (116) from historical query terms (116) mined from a query log (118);

sending respective ones of at least a subset of the low FOO historical query terms (116) to the search engine (126) to obtain search results;

extracting features from at least a subset of search results; and

producing the term vectors (136) from the features as a function of term (108) and inverted term frequencies; and

further comprising after clustering:

determining that there is no match between the term (108) and term(s) from the first set of term clusters (138), the first set being based on high FOO historical query terms (116); and

responsive to the determining, identifying a match between the term (108) and term(s) from one or more of
the second set of term clusters (138), the second set being based on low FOO historical query terms (116); and

responsive to identifying, generating related term suggestion(s) comprising the term(s)."

XII. Claim 1 of auxiliary request III differs from claim 1 of auxiliary request II in that after "producing the term vectors ... inverted term frequencies" the following text has been inserted:

"wherein the making the second set of term clusters is performed using a trained classifier (140), generated from high FOO query terms;"

XIII. Claim 1 of auxiliary request IV differs from claim 1 of auxiliary request I in that the following text has been added at the end of the claim:

"otherwise,

    generating a trained classifier (140) from the term clusters (138) which are based on high FOO query terms (120) by using a statistical classification and machine learning technique;

    sending low FOO query terms (122), one by one, to the search engine (132) and receiving corresponding search results (130);

    extracting extracted features (134) from the search results (132), and generating term vectors (136) therefrom;

    classifying the term vectors (136) generated from
the low FOO query terms (122) in view of the trained classifier (140) to generate respective term clusters (138) based on the low FOO query terms (122);

generating a suggested term list (110) from the terms from the term clusters (138) based on the low FOO query terms (122) that are determined to be substantially similar to the term (108); and

sending the suggested term list (110) to the entity."

XIV. The appellant's arguments as relevant to this decision are discussed in detail below.

Reasons for the Decision

1. The appeal complies with the provisions referred to in Rule 101 EPC and is therefore admissible.

2. The invention

2.1 The background section of the description highlights the importance of identifying popular keywords relevant to a website, in particular in the context of search engine result optimisation (paragraphs [0003] and [0004] of the description of the application as filed). To allow relevant keywords to be identified without human intervention, the application proposes an automated method for "related term suggestion".

2.2 This method starts with a set of "historical query terms" obtained from query logs. Given a "bid term", it produces one or more ordered lists of query terms
related to the bid term (paragraphs [0006], [0039] and [0040]; Figure 2). The terms in each list are meant to relate to a particular "sense" of the bid term, corresponding to the context in which the bid term is used (paragraphs [0015] and [0032]). One example is given in Table 1 on page 8 of the application as filed, which shows that the "mail" bid term results in one list of query terms related to the context "online e-mail" and another list of query terms related to the context "traditional mail" (paragraphs [0021] to [0023]).

2.3 The method first divides the historical query terms into a set of query terms having a high frequency of occurrence (FOO) in the historical query logs and a set of query terms having a low FOO. A historical query term is a "high FOO" query term if the number of times it occurs is greater than or equal to a configurable threshold value (paragraph [0025]). The high-FOO query terms are then assigned to "term clusters" on the basis of their mutual "similarities" (paragraph [0040]; paragraphs [0030] and [0031]).

2.4 The "similarity" of two high-FOO query terms is determined as follows. For each query term, a "term vector" is constructed from a textual search result associated with the query term (paragraphs [0025] to [0028]). A term vector is a vector whose dimensions correspond to keywords occurring in the search result. The value of a dimension is the "term frequency-inverse document frequency" (tf-idf) score of the corresponding keyword, which is a measure of the FOO of the keyword in the search result (paragraph [0029]). The similarity between two query terms is then determined by calculating the inner product of their (normalised) term vectors, which is equivalent to calculating the
cosine of the angle between the two vectors (cf. paragraph [0030]). Effectively, this gives a measure of the textual similarity of the corresponding search results.

2.5 According to paragraph [0031], the generation of "term clusters" of "similar" query terms is performed by means of the known "density-based clustering algorithm" (DBSCAN).

2.6 As explained in paragraph [0032], after generation of the term clusters, a bid term received from an "entity" (such as a user) is "matched" with the query terms in the various term clusters. This matching operation involves a string-based comparison and looks for an exact or partial match. For each term cluster containing a query term that matches the bid term, a "term list" is returned consisting of the query terms in the term cluster ordered "by a combination of FOO and confidence value", where the confidence value is based on the "similarity" between the query term and the bid term. In case of matches with multiple term clusters, the returned term lists are ordered by term cluster size (paragraph [0033]).

2.7 The procedure may be extended to low-FOO query terms if no match is found between the received bid term and a high-FOO query term (paragraphs [0040] and [0041]; Figures 2 and 3). In this case, search results and corresponding term vectors are also generated for the low-FOO query terms, and expanded term clusters are constructed by means of a "classifier" trained on the term clusters generated from the high-FOO query terms (paragraphs [0034] to [0037]). The bid term is then matched against the expanded term clusters (paragraph [0038]).
3. **Main request - inventive step**

3.1 Claim 1 defines a computer-implemented method for "related term suggestion" as described in points 2.2 to 2.6 above.

Apart from the statement that the method is "computer-implemented", the only further reference to technical means occurs in the feature "each vector being generated from search results associated with the term which is a term of a set of high FO0 historical query terms previously submitted to a search engine". The Board considers, however, that the words "previously submitted" do not express the submission of query terms to a search engine as a step of the claimed method but - in so far as they limit the claimed subject-matter - only characterise the informational content of the query terms and, arguably, the search results.

The remaining features of claim 1 express the algorithm underlying the computer-implemented method in terms of linguistic and statistical functional features. Such features, which "as such" fall under the exclusions of Article 52(2) and (3) EPC, provide a technical contribution only to the extent that they interact with the technical subject-matter of the claim for solving a technical problem (see decision T 154/04, OJ EPO 2008, 46, reasons 5, under (F), and reasons 13).

3.2 The algorithm underlying claim 1 serves the overall purpose of suggesting query terms that are semantically related to the various "senses" of a particular input term. This is not a technical problem, for whether terms are "related" to each other is a cognitive or linguistic matter and not a technical issue (cf.
decisions T 1358/09 of 21 November 2014, reasons 5.2; T 2230/10 of 3 July 2015, reasons 3.9; and T 2439/11 of 11 November 2016, reasons 4.5).

Nevertheless, a technical interaction may be present if technical considerations motivating the algorithm's design can be identified that make the algorithm particularly suitable for being performed on a computer and that "go beyond merely finding a computer algorithm to carry out some procedure" (cf. decision T 1358/09, reasons 5.5).

3.3 In the statement of grounds of appeal, the appellant submitted that the claimed ordering of terms within a suggested-term list based on "confidence value" involved such technical considerations, because it re-used the similarity measure that had been calculated in a previous method step. Those technical considerations allowed an algorithm for related-term suggestion with reduced computational effort to be implemented.

In the Board's view, however, the consideration that an intermediate result produced by an earlier algorithmic step may be re-used in a later step is an algorithmic rather than a technical consideration, as it does not require considerations about the internal functioning of a computer. This is in line with the case law of the boards of appeal, which generally holds that algorithmic efficiency is not a technical effect (cf. decisions T 1784/06 of 21 September 2012, reasons 3.1.2; T 42/10 of 28 February 2013, reasons 2.11; T 1370/11 of 11 March 2016, reasons 10 to 10.5).

3.4 At the oral proceedings, the appellant argued that the claimed use of a "configurable threshold value" was a
technical aspect, because it made the method more flexible and allowed its use in more circumstances.

The Board considers the use of a threshold value for distinguishing between high-FOO and low-FOO query terms to be a purely algorithmic choice and therefore not a technical aspect. Making this value "configurable" arguably is technical, at least to the extent that it implies the presence of an input mechanism allowing the user to configure the value. But parameter configuration and its advantages are well known in the art.

3.5 The appellant also argued that finding related terms by "matching" the bid term with the query terms in the term cluster was technical, because matching strings was computationally efficient.

The claim is silent, however, on the technical implementation of the matching operation. The argument therefore again relies on the claimed method being algorithmically efficient, which is not a technical effect.

3.6 The Board therefore concludes that the algorithm underlying the computer-implemented method of claim 1 provides no technical contribution. Implementing the algorithm on a computer, including a facility for configuring the threshold value, is a routine exercise for the skilled person; in fact, the claim merely recites the term "computer-implemented", and the description contains no technical details of the software implementation.

3.7 As explained in point 3.1 above, the Board does not interpret the expression "a set of high FOO query terms
previously submitted to a search engine" as a step of the claimed method. In any event, once the non-technical decision has been made to base the related-term suggestion algorithm on the textual content of search results associated with historical query terms, it is obvious to obtain such search results by submitting the terms to a search engine.

3.8 The subject-matter of claim 1 of the main request therefore lacks inventive step over a notorious general-purpose computer (Article 56 EPC).

3.9 The Board notes that the Examining Division took document D5 as closest prior art. Document D5 relates to a method for automatically generating a "query taxonomy". Like the method of the present application, it starts with a set of queries \( q \), constructs associated term vectors \( v_q \) by means of conventional tf-idf techniques from "highly ranked search-result snippets" returned by Web search engines or information retrieval systems and takes the cosine of the angle between term vectors as a measure of the similarity between the corresponding queries (see section 3). The queries are then classified into clusters on the basis of their mutual similarities (section 4). Term suggestion is mentioned as one of the possible applications of the query taxonomy thus generated (section 7).

3.10 Although the teaching of document D5 in terms of purpose and common features may be "closer" to the claimed invention than a notorious general purpose computer, this does not invalidate the Board's finding of lack of inventive step over a notorious general purpose computer: Article 56 EPC requires the claimed invention to be non-obvious having regard to the state
of the art. Thus, there is no need for the Board to consider whether the same conclusion of lack of inventive step would be reached when starting from document D5 (see e.g. decisions T 1928/06 of 20 October 2009, reasons 1.3.2; T 1734/11 of 13 January 2015, reasons 5.4; and T 1742/12 of 22 June 2016, reasons 6.6 and 10.3).

4. **Auxiliary request I - inventive step**

Claim 1 of auxiliary request I differs from claim 1 of the main request in a number of clarifications. Since these clarifications do not affect the inventive-step reasoning given above, the subject-matter of claim 1 of auxiliary request I likewise lacks inventive step (Article 56 EPC).

5. **Auxiliary requests II and III - inventive step**

5.1 The features added to claim 1 of each of auxiliary requests II and III extend the method for related-term suggestion to low-FOO query terms, as described in point 2.7 above.

Claim 1 of auxiliary request III is more precise, in that the added feature "wherein the making the second set of term clusters is performed using a trained classifier, generated from high FOO query terms" clarifies that the classification of the low-FOO query terms into term clusters is based on the clusters formed from high-FOO query terms.

5.2 The Board observes that claim 1 of both requests refers to a "second set" of term clusters to be generated on the basis of the low-FOO historical query terms, suggesting that the low-FOO query terms are clustered
into a set of clusters separate from the clusters containing the high-FOO query terms, and not into "expanded term clusters", as the teaching disclosed in paragraph [0038] does. At the oral proceedings, the appellant submitted that the words "second set" may be understood as distinguishing the original (first) set of unexpanded term clusters from the set of expanded term clusters, i.e. the original term clusters but now including the low-FOO query terms.

The Board accepts this interpretation for the purpose of assessing inventive step in respect of auxiliary requests II and III.

5.3 The method according to claim 1 of auxiliary requests II and III hence implements a two-step term clustering approach. In the first step, the high-FOO query terms are taken all together and divided into clusters by means of a clustering algorithm. In the second step, the remaining low-FOO query terms are added one by one to the clusters formed in the first step by means of a "trained classifier". The appellant argued that classifying all query terms using this two-step approach was computationally more efficient than classifying them all in a single step, which would have been the approach taken by the skilled person.

However, the claimed two-step approach still concerns the algorithm and is, in the Board's view, not based on technical considerations going beyond "merely finding a computer algorithm". That the approach may be more efficient than other approaches (which would however depend on the specific clustering algorithm used and on its implementation), and also less straightforward, would be relevant if the algorithm were technical. But the Board considers that it is not.
The definition of the second step explicitly states that the search results from which the term vectors for the low-FOO query terms are generated are obtained by "sending respective ones of at least the subset of the low FOO historical query terms to the search engine".

Submitting a query term to a search engine to obtain search results is technical, but it is also obvious once the non-technical decision has been made to base the related-term suggestion algorithm on the textual content of search results associated with historical query terms (cf. point 3.7 above).

The final aspect of claim 1 of both requests to be considered is the decision to add the low-FOO query terms to the term clusters formed from high-FOO query terms only if no match is found between the received bid term and the high-FOO query terms. Since it is reasonable to assume that the term clusters generated by the method are not used only once for a single bid term but are used to find related terms for multiple received bid terms, this decision has the effect that the returned results do not include low-FOO query terms until the first bid term is received that does not match any of the high-FOO query terms. Since the Board cannot see any technical effect in this, and the appellant has not alleged any, this aspect, too, cannot support an inventive step.

Hence, the subject-matter of claim 1 of auxiliary requests II and III lacks inventive step (Article 56 EPC).
6. **Auxiliary request IV - inventive step**

6.1 Claim 1 of auxiliary request IV, despite differences in wording, largely corresponds to claim 1 of auxiliary request III. It adds that the "trained classifier" is generated by means of a "statistical classification and machine learning technique". It refers to "term clusters based on the low FOO query terms" rather than a "second set of term clusters", but the Board agrees with the appellant that this may still be interpreted as referring to the "expanded term clusters" discussed in paragraph [0038] of the description (cf. point 5.2 above).

6.2 The reference to a "statistical classification and machine learning technique" limits the choice of algorithm used for assigning low-FOO query terms to an appropriate term cluster, but this algorithmic feature does not render the non-technical algorithm technical.

6.3 The Board therefore concludes that the subject-matter of claim 1 of auxiliary request IV also lacks inventive step (Article 56 EPC).

7. **Conclusion**

Since none of the appellant's substantive requests is allowable, the appeal is to be dismissed.
Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:  The Chairman:

I. Aperribay  R. Moufang

Decision electronically authenticated