Internal distribution code:

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

Datasheet for the decision of 6 November 2014

Case Number: T 0598/14 - 3.5.07

Application Number: 04768572.2

Publication Number: 1668541

IPC: G06F17/30

Language of the proceedings: EN

Title of invention:
Information retrieval

Applicant:
British Telecommunications plc

Headword:
Enhanced retrieval/BRITISH TELECOMMUNICATIONS

Relevant legal provisions:
EPC Art. 56, 112(1)(a)

Keyword:
Inventive step - mixture of technical and non-technical features (distinguishing features non-technical)
Referral to the Enlarged Board of Appeal - (no)

Decisions cited:
G 0003/98, G 0003/08, G 0001/12, T 0162/90, T 1177/97, T 0520/01, T 0154/04, T 1351/04, T 1569/05, T 1875/07

Catchword:

Europäisches Patentamt
European Patent Office
Office européen des brevets

Beschwerdekammern
Boards of Appeal
Chambres de recours

European Patent Office
D-80298 MUNICH
GERMANY
Tel. +49 (0) 89 2399-0
Fax +49 (0) 89 2399-4465

Case Number: T 0598/14 - 3.5.07

D E C I S I O N
of Technical Board of Appeal 3.5.07
of 6 November 2014

Appellant: (Applicant)	British Telecommunications plc 81 Newgate Street London EC1A 7AJ (GB)
Representative:	Lau, Chi-Fai British Telecommunications plc Intellectual Property Department, PPC5A BT Centre 81 Newgate Street London EC1A 7AJ (GB)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 23 October 2013 refusing European patent application No. 04768572.2 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	R. Moufang
Members:	P. San-Bento Furtado
	R. de Man

Summary of Facts and Submissions

I. The appeal lies from the decision of the Examining Division to refuse European patent application No. 04768572.2, which originated from the international application published as WO 2005/041063. The application concerns an information retrieval apparatus for enhanced retrieval of documents based on semantic similarity of search terms.

II. In the contested decision, reference was made to document D11:

D11: Baeza-Yates R., Ribeiro-Neto B., "Modern Information Retrieval", "Chapter 2: Modeling", Addison-Wesley, Harlow, GB, 1999.

The Examining Division decided that the subject-matter of independent claim 1 lacked an inventive step over the disclosure of document D11.

III. In the statement of grounds of appeal, the appellant requested that the decision be set aside and that a patent be granted on the basis of the claims on which the decision was based, filed during oral proceedings on 10 October 2013.

IV. In a communication accompanying a summons to oral proceedings, the Board expressed its preliminary view that the subject-matter of independent claim 3 did not involve an inventive step over prior-art document D11, nor did the subject-matter of the other independent claims. Furthermore, independent claims 3 and 6 did not appear to fulfil the requirements of clarity and conciseness.

V. With a letter dated 6 October 2014, the appellant filed arguments in response to the preliminary view of the Board.

VI. During the oral proceedings held on 6 November 2014 the appellant initially replaced its pending request by a new request in which the previous claims 6 and 7 were deleted. Later the appellant replaced this request with a new main request filed at 9.35 hrs comprising claims 1 to 3, corresponding to previous claims 3 to 5. Furthermore, the appellant requested the Board to refer the following question to the Enlarged Board of Appeal:

> "Does the use of a mathematical method, namely the determination of probabilities indicating a degree of semantic similarity, as one step of a method claim comprising multiple steps and having an acknowledged further technical effect render the method claim unpatentable?"

At the end of the oral proceedings, the chairman pronounced the Board's decision.

VII. The appellant's final substantive requests were that the decision under appeal be set aside and that a patent be granted on the basis of the main request filed at the oral proceedings. The appellant also made a final procedural request that a question of law be referred to the Enlarged Board of Appeal.

VIII. Claim 1 of the sole request reads as follows:

"An information retrieval apparatus for use in retrieving information from a set of one or more documents, comprising:
an input for receiving a search query;

generating means for generating a set of probabilities indicative of the semantic similarity of words selected from said set of one or more documents;

query enhancement means for modifying a received search query with reference, in use, to said generated set of probabilities; and

information retrieval means for searching said set of one or more documents for relevant information using a received search query modified by said query enhancement means,

wherein said generating means are arranged, in use:

(i) for each word selected from said set of one or more documents:

(a) to identify, in documents of said set of one or more documents, word sequences comprising the word and a predetermined number of other words;

(b) to calculate a relative frequency of occurrence for each distinct word sequence among word sequences containing the word; and

(c) to generate a fuzzy set comprising, for groups of word sequences containing the word, corresponding fuzzy membership values calculated from the relative frequencies determined at step (b); and

(ii) to calculate, for each pair of words of said plurality of words, using respective fuzzy sets generated at step (i), a probability that the first word of the pair is semantically suitable as a replacement for the second word of the pair."

IX. The reasons given in the decision under appeal relating to the then sole request can be summarised as follows:

Document D11 disclosed a method for determining the semantic similarity of words in a plurality of words selected from a set of one or more documents, for use in the retrieval of information in an information system comprising some of the features of the invention. Prior-art document D11 also disclosed the use of fuzzy sets for query expansion.

The claimed method differed from the method of document D11 only as regarded the underlying mathematical fuzzy model for calculating the fuzzy sets and the semantic similarity. This difference, however, related to non-technical aspects. Following the approach of the Guidelines for Examination in the EPO, G-VII, 5.4, the objective problem could be established as how to implement the fuzzy thesaurus based on the mathematical fuzzy model. That implementation was straightforward for the skilled person.

The Examining Division considered information retrieval per se not to be technical; only its technical implementation in a computer system made it technical.

X. In the appeal proceedings, the appellant argued essentially as follows:

The Examining Division had failed to properly interpret and apply the established case law.

The search according to the invention was not a search for a single word but for a word sequence. Therefore, the search was not the same as in the method disclosed in document D11. This distinguished the present case from that of decision T 1569/05. Whereas T 1569/05 related to a search performed within a restricted search space, with the search being carried out in a

conventional manner, according to the current invention a different enhanced search was performed.

The invention returned, for the same search string, a different set of documents. The search result was better. These were further technical effects.

Furthermore, decision T 1569/05 stated in point 3.7 of its reasons that the case at hand was to be distinguished from the subject-matter considered in decision T 1351/04, where the Board had seen a technical effect in "the control of the computer along the path leading to the desired data" (point 7.2). It was believed that the present application also related to the control of the computer in searching for specific word sequences.

Regarding the referral of a question of law to the Enlarged Board of Appeal, the appellant argued that a restrictive interpretation of the law would make it almost impossible to get patents in the area of information retrieval. However, information retrieval systems brought significant technical advantages and were a very important area of great technological progress.

Reasons for the Decision

The invention

1. The application relates to a method for generating, from an input set of documents, a word replaceability matrix defining semantic similarity between words occurring in the input document set. The word replaceability matrix can be used for determining

document similarity and to enhance search queries for retrieval of information from the document set (page 4, line 31, to page 5, line 4, of the international publication of the application as filed).

According to the method of the invention, for each word distinct word sequences of a predetermined length are identified from the documents of the input set. Each word sequence comprises the word and gives an indication of the context in which it was used.

The relative frequencies of occurrence of the identified word sequences for the word are taken into account to calculate fuzzy sets for each word comprising membership values for corresponding groups of word sequences. For each pair of words occurring in the document set, their respective fuzzy sets are used to calculate the probability that the first word of a pair is semantically suitable as a replacement for the second word of the pair, these probabilities being collated to form the word replaceability matrix.

The word replaceability matrix may be used to improve document clustering or to extend or modify the set of search words in a user's search query.

Articles 52(1) and 56 EPC

2. The Board is of the opinion that the Examining Division correctly applied the established case law and rightly concluded that the claimed subject-matter was not inventive.

2.1 Document D11 consists of passages from a text book on information retrieval. In the introduction of chapter 2, on pages 19 to 21, an information retrieval

system is described as a system for retrieving information in the form of relevant documents from a set of one or more documents in accordance with a user query. Information systems are said to usually adopt index terms to index and retrieve documents, an index term being a keyword (or group of related words) which has some meaning of its own, or simply any word appearing in a document. Retrieval is based on index terms.

Such a system therefore comprises "an input for receiving a search query", for example in the form of terms, query processing means for treating this query, and "information retrieval means for searching a set of one or more documents for relevant information using a received search query". The Board considers these features to define a standard information retrieval system for retrieval of documents, characterised in terms of its basic components.

Document D11 then describes that in order to predict which documents are relevant a ranking algorithm is used which is based on premises regarding document relevance. Distinct sets of premises yield different information retrieval models (page 19, last paragraph). Page 20, third full text paragraph, of document D11 mentions that information retrieval systems may be based on fuzzy models, which are further described on page 34, last paragraph and following pages.

According to the passage on page 35, last paragraph, to page 36, third paragraph, a thesaurus can be used to expand the set of index terms in the query with related terms (obtained from the thesaurus), or to "model the information retrieval problem in terms of fuzzy sets". Document D11 also explains that a thesaurus can be

built by defining a "term-term correlation matrix", each entry in the matrix indicating a correlation factor between two index terms in the document collection.

Therefore, the skilled person reads from that passage of document D11 that an information retrieval system may also comprise "generating means for generating a set of probabilities indicative of the semantic similarity of words selected from the set of documents".

The means for expanding the set of terms in the query described in the first lines of page 36 correspond to "query enhancement means for modifying a received search query with reference, in use, to said generated set of probabilities". This latter feature may be interpreted, in the light of the description, for example page 13, lines 15 to 20, and original claim 4, as encompassing using said probabilities to identify words with similar meaning to those of a term of the received search query, and modifying the received search query to also include the identified words as search terms.

The correlation matrix of document D11 is calculated for the index terms of the set of documents (page 36, second full paragraph). Since index terms can be any words in the documents (page 19 of document D11), that calculation corresponds to calculating "for each pair of words of said plurality of words" "a probability that the first word of the pair is semantically suitable as a replacement for the second word of the pair".

Document D11 also discloses on page 36 the use of the correlation matrix to calculate fuzzy sets associated to the index terms. The degree of membership of a document in a fuzzy set is calculated based on the values of the correlation matrix. However, these fuzzy sets are different from those of the claim and are not used for generating a thesaurus.

2.2 The information retrieval apparatus of the claim therefore differs from that of document D11 in that the calculation of the probability that a first word of a pair is semantically suitable as a replacement for the second word of the pair is done by: identifying word sequences comprising the word, calculating relative frequencies of occurrence for distinct word sequences, and generating fuzzy sets comprising corresponding fuzzy membership values calculated from the relative frequencies (features (i)(a) to (i)(c)). Furthermore, the fuzzy sets are used to calculate the probability (feature (ii)).

2.3 The Board is of the opinion that following established case law these distinguishing features have to be considered non-technical aspects which cannot contribute to the inventive step of the claimed apparatus.

The distinguishing features correspond to a change in the mathematical model used for calculating the probability that a first word of a pair is semantically suitable as a replacement for the second word of the pair. The mathematical model of the claims is based on determining fuzzy sets by taking into account, for each word or term, the word sequences with a predetermined number of words of the context of the term.

As argued by the Examining Division, the underlying considerations are which terms have to be taken into account for defining semantic similarity. On page 5, lines 10 to 12, of the description it is said that the main purpose of identifying *n*-grams (sequences of *n* words) was to understand and represent the context in which the particular words were used in a document. Pages 7 to 10 give examples of sentences in which the word "brown" occurs and how the occurrences of the terms in 3-grams is used to generate fuzzy sets and calculate the probability that one word "may be a semantically suitable word to use in place of" another word.

The considerations are therefore that, in text documents, words which have similar meanings or are related are more likely to occur in the same or similar phrases or context than unrelated words. These considerations are of a purely linguistic nature. According to decision T 1177/97, "Translating natural languages/SYSTRAN", of 9 July 2002, point 3, last paragraph, "Features or aspects of the method which reflect only peculiarities of the field of linguistics, however, must be ignored in assessing inventive step".

In its letter dated 6 October 2014 the appellant argued that the present invention did not rely on any linguistic features because the words of the documents were not "analysed to determine any common root or features" and that there was no "attempt to determine the semantic meaning or context of any particular word". It is true that the invention does not directly model the semantic meaning of the words but it does examine the context of words in order to find, by statistical analysis, words which are "semantically

suitable" to use in place of other words. These aspects are linguistic issues.

In the invention, the linguistic aspects are translated into the mathematical model. The Board considers that the translation of linguistic considerations into a mathematical model with the aim of enabling the linguistic analysis to be done automatically by a computer can be seen as involving, at least implicitly, technical considerations. This is also in line with decision T 1177/97, point 3, last paragraph, or opinion G 3/08, "Programs for computers", OJ EPO 2011, 10, points 13.2 and 13.3. However, according to G 3/08, point 13.5, this is not enough to guarantee the technical character of subject-matter otherwise excluded from patentability under Article 52(2) and (3) EPC. The technical character would have to be established on the basis that those considerations constituted "further technical considerations".

The Board is convinced that no such "further technical considerations" can be found in the present case. As explained above, the translation simply reflects the linguistic aspects in the mathematical model. The modified model for semantic similarity results in a different set of words being considered to be semantically suitable as a replacement for each word. According to established case law, such linguistic aspects do not have a technical character.

2.4 The appellant argued that the present invention, in particular the analysis of a number of word sequences, had the advantage of supporting the generation of asymmetric relationships between the terms. For example, as described on page 10, lines 11 to 15, the probability that "black" was a suitable replacement for

"brown" was not the same as the probability that "brown" was a suitable replacement for "black". Such an asymmetric relationship could not be derived from any prior art.

The Board recognises this difference between the described invention and the apparatus of document D11, but does not consider that it contributes to a further technical effect in accordance with established case law. The different search results are a consequence of the non-technical linguistic aspects taken into account.

The appellant also argued that query enhancement, or producing different results for the same query, or obtaining a better result, were further technical effects. The Board cannot agree. Query enhancement is also done in the apparatus of document D11, in which search queries are modified to also include semantically similar terms. The search results produced by the claimed apparatus are distinct from those of the prior art essentially in that information with a different semantic content is retrieved. This is a non-technical distinction, not a further technical effect. Furthermore, the concept of "better search" is subjective in the context of retrieval based on semantic similarity.

2.5 In the grounds of appeal the appellant argued that the present case was different from that of decision T 1569/05, "Method for retrieving data/CANON" of 26 June 2008, which related to a search performed within a restricted search space, with the search being carried out in a conventional manner. In the present invention a different search was performed. The present application also related to the control of the computer

in searching for specific word sequences, as in decision T 1351/04, "File search method/FUJITSU" of 18 April 2007.

The Board agrees that the present invention is different from the method for retrieving data discussed in T 1569/05. However, as in T 1569/05, point 3.2, the difference to the prior art lies in a mathematical model used for data retrieval. Similarly to that case, the present application deals with information retrieval at an abstract level. It is clear for the skilled person that in the claimed information retrieval apparatus the information is stored as data in physical locations of the computer memory. However, the present application does not describe how the computer finds the stored data corresponding to the information to be retrieved. As in the case considered in T 1569/05, point 3.7, these technical aspects are not part of the invention.

This is quite distinct from the case of T 1351/04, where the invention relates to an index containing "management information" which "controls the computer by directing it to a certain memory location" (point 7.2). The index of T 1351/04 is a data structure including information about the position of records in a file. The technical effect in T 1351/04 is considered to be "the control of the computer along the path leading to the desired data" (point 7.2). In the present case, such an effect is absent because the distinguishing features are not concerned with such control.

2.6 In this regard, the Board notes that the "index terms" mentioned in document D11, page 19, first paragraph, often referred to in the context of information

retrieval systems such as those of document D11 and the present application, are to be distinguished from the index discussed in T 1351/04, which is a data structure to access data in memory.

As defined in document D11, an "index term" is a keyword which has some meaning of its own or simply any word which appears in the text of a document in the collection. The designation "index term" therefore often refers to a non-technical abstract concept.

Index structures are data structures keeping information about the location of data stored in memory for the purpose of supporting access to data, in particular, for efficient retrieval of data.

Index structures, for example trees or hash lists, are required in file systems, database systems or information retrieval systems such as the present one. However, the application does not describe any index structures used in the information retrieval system of the invention.

2.7 For the above reasons, the Board concludes that the subject-matter of independent claim 1 lacks an inventive step (Articles 52(1) and 56 EPC).

Request for referral of a question to the Enlarged Board

3. In oral proceedings the appellant requested the referral of the following question to the Enlarged Board of Appeal:

"Does the use of a mathematical method, namely the determination of probabilities indicating a degree of semantic similarity, as one step of a method claim

comprising multiple steps and having an acknowledged further technical effect[,] render the method claim unpatentable?"

3.1 According to Article 112(1)(a) EPC, the Board should refer a question to the Enlarged Board of Appeal if it considers that a ruling is required to ensure uniform application of the law or if a point of law of fundamental importance arises. However, particularly in the field of new technologies, it is primarily the duty of the Technical Boards of Appeal to interpret and apply the European Patent Convention (G 3/08, reasons 7.3.2 to 7.3.5, T 0154/04, OJ EPO 2008, 46, "Estimating sales activity/DUNS LICENSING ASSOCIATES, L. P.", reasons 2 and 4).

3.2 The appellant argued that the area of information retrieval had made a valuable contribution to technological progress in the recent past but that it was almost impossible to obtain patents in the field. For recognising technical character the law was applied very restrictively. The question was therefore of fundamental importance.

3.3 The Board accepts that the question is relevant for a significant number of cases, a criterion to take into account when judging whether a point of law is of fundamental importance (G 1/12 of 30 April 2014, reasons 10).

3.4 However, the established case law already gives a clear answer to the question. In particular, it is established that any claim involving technical means, or a further technical effect, is not excluded from patentability by Article 52(2) and (3) EPC (G 3/08, reasons 10.7 and 10.7.1). Therefore, if the method has

an acknowledged further technical effect it will not be rendered unpatentable, in the sense of Article 52(2) and (3) EPC, by the use of a mathematical method. Therefore, the referral of the question is not necessary to ensure uniform application of the law.

3.5 In addition, it has also been established in the case law that an answer to the question to be referred must be necessary for reaching a decision on the appeal (see for example G 3/98, "Six-month period/UNIVERSITY PATENTS", OJ EPO 2001, 62, reasons 1, T 520/01 of 29 October 2003, reasons 4.1, and T 1875/07 of 5 November 2008, reasons 7).

A referral is therefore ruled out if the question does not arise in the case under appeal, for example because it is based on assumptions which do not apply to the case (see also T 162/90 of 7 May 1992, reasons 7).

In the present case the proposed question does not need to be answered because it relies on incorrect assumptions. The question relates to a method claim whereas the decision is based on the apparatus claims submitted by the appellant as a final request. Furthermore, the Board does not argue that the "use of the mathematical method" renders the claimed subject-matter unpatentable, as mentioned in the question, but rather that the difference of the claimed invention to the closest prior art does not have a further technical effect and does not contribute to an inventive step.

3.6 In the opinion of the Board it would not be appropriate to refer an improved, reformulated question either. As explained above, it is primarily the task of the Technical Boards of Appeal to develop the case law in new technology fields. In this role, the Board finds

that a referral is not justified by either the particulars of the present case nor the uniform application of the law.

As shown by the discussion above, in the Board's view there is clear established case law on the question of patentability of mathematical methods or methods for information retrieval based on semantic similarity, and the application of that case law is harmonised.

The Board further notes that the undisputable importance of information retrieval in contemporary industry does not mean that all relevant developments in the area should be considered patentable. Similarly, the fact that specific subject-matter excluded from patentability, for example a discovery or a scientific theory, is very relevant for an important sector of industry is not decisive in applying the criteria of Article 52(2) and (3) EPC. As a side-note the Board remarks that there are many technical inventions not excluded from patentability in the field of information retrieval (see also point 2.6 above), so that it cannot follow the appellant's assertion that it is almost impossible to obtain patents in the area.

3.7 The request for referral of a question of law to the Enlarged Board of Appeal is therefore refused under Article 112(1)(a) EPC.

Order

For these reasons it is decided that:

1. The request to refer a question of law to the Enlarged Board of Appeal is refused.

2. The appeal is dismissed.

The Registrar: The Chairman:

I. Aperribay R. Moufang

Decision electronically authenticated