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

Datasheet for the decision
of 21 November 2014

Case Number: T 1358/09 - 3.5.07

Application Number: 00926837.6

Publication Number: 1175652

IPC: G06F17/30

Language of the proceedings: EN

Title of invention:
Classification method and apparatus

Applicant:
BDGB Enterprise Software Sàrl

Headword:
Classification/BDGB ENTERPRISE SOFTWARE

Relevant legal provisions:
EPC Art. 56

Keyword:
Inventive step - (no)

Decisions cited:
G 0003/08, T 0258/03, T 0154/04, T 1784/06, T 1954/08,
T 1316/09

Catchword:

Case Number: T 1358/09 - 3.5.07

D E C I S I O N
of Technical Board of Appeal 3.5.07
of 21 November 2014

Appellant: (Applicant)	BDGB Enterprise Software Sàrl Av. Du Vignoble 2 3rd Floor West 2009 Neuchâtel (CH)
Representative:	Jordan, Volker Otto Wilhelm Weickmann & Weickmann Patentanwälte Postfach 860 820 81635 München (DE)
Decision under appeal:	Decision of the Examining Division of the European Patent Office posted on 12 February 2009 refusing European patent application No. 00926837.6 pursuant to Article 97(2) EPC.

Composition of the Board:

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

Summary of Facts and Submissions

I.   The applicant (appellant), which at the time was SER
        Systeme AG Produkte und Anwendungen der
        Datenverarbeitung, lodged an appeal against the
        decision of the Examining Division refusing European
        patent application No. 00926837.6.

II.   With effect from 25 March 2011, the application was
        transferred to BDGB Enterprise Software Sàrl, which
        thereby obtained the status of appellant.

III.   According to the contested decision, the application
        did not comply with Article 83 EPC and the subject-
        matter of claim 1 lacked an inventive step within the
        meaning of Article 56 EPC. No documents were cited.

IV.   With the statement of grounds of appeal, the appellant
        essentially requested that the decision under appeal be
        set aside and that a patent be granted on the basis of
        the claims filed with the letter dated 5 June 2007,
        i.e. the claims on which the decision under appeal was
        based.

V.   In a communication accompanying a summons to oral
        proceedings, the Board indicated that, although it
        could not agree with the reasons given in the contested
        decision, the application appeared not to sufficiently
        disclose the invention over its whole claimed scope.
        Assuming the appellant succeeded in overcoming this
        objection, the subject-matter of claim 1 appeared to
        lack an inventive step.

VI.   With a letter dated 21 October 2014, the appellant
        informed the Board that it had decided "to abandon this
        case" and so would not attend the oral proceedings. In

addition, it requested a decision according to the
state of the file.

VII.   Oral proceedings took place on 21 November 2014 in the
        absence of the appellant. At the end of the oral
        proceedings, the chairman pronounced the Board's
        decision.

VIII.   Claim 1 reads as follows:

        "A method for the computerized classification of an
        unclassified text document into one of a plurality of
        predefined classes based on a classification model
        obtained from the classification of a plurality of
        preclassified text documents which respectively have
        been classified as belonging to one of said plurality
        of classes, said document and said documents
        respectively comprising a plurality of terms which
        respectively comprise one or more symbols of a finite
        set of symbols;

        a)   wherein said method involves the computerized
             building of said classification model, comprising
             the following method steps:
             a1)   representing each of said plurality of text
                   documents, which are digitally represented
                   in a computer, by a vector of n dimensions,
                   said n dimensions forming a vector space,
                   whereas the value of each dimension of said
                   vector corresponds to the frequency of
                   occurrence of a certain term in the document
                   corresponding to said vector, so that said n
                   dimensions span up a vector space;
             a2)   representing the classification of said
                   already classified documents into classes by
                   separating said vector space into a

plurality of subspaces by calculating one or more hyperplanes, such that each subspace comprises one or more documents as represented by their corresponding vectors in said vector space, so that said each subspace corresponds to a respective class;

    a3)   calculating a maximum margin surrounding said hyperplanes in said vector space such that said margin contains none of the vectors contained in the subspaces corresponding to said classification classes;

b)   wherein said method further involves, on basis of said classification model, the computerized classification of said unclassified text document as belonging to one of said plurality of classes, comprising the following method steps:

    b1)   representing said text document, which is digitally represented in a computer, by a vector of n dimensions, said n dimensions spanning up said vector space, whereas the value of each dimension of said vector corresponds to the frequency of occurrence of a certain term in the document corresponding to said vector;

    b2)   classifying said document into one of said plurality of classes by determining into which of said plurality of subspaces of said vector space said vector falls and identifying said document as belonging to a certain class which corresponds to the subspace into which said vector falls;

    b3)   calculating a confidence level for the classification of said document as belonging to said certain class based on the distances between the vector representing said

document and all hyperplanes surrounding
said subspace which corresponds to said
certain class normalized by the
corresponding margins such that a document
which lies outside said margins is assigned
a confidence level of '1' and a document
which falls into said margins is assigned a
value between '0' and '1'."

Reasons for the Decision

1.   The appellant's declaration made in the letter dated
        21 October 2014 that it had decided "to abandon this
        case" in itself does not amount to an unambiguous
        declaration of the withdrawal of the appeal. Since in
        the same letter the appellant requested a decision
        according to the state of the file, there is in fact no
        doubt that it intended not to withdraw the appeal, but
        to inform the Board that it would make no further
        submissions. The Board is therefore competent to decide
        on the appeal.

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

3.   *The invention*

3.1   The invention is concerned with the computerised
        classification of text documents. This is done by first
        building a "classification model" and then classifying
        documents using this classification model.

3.2   The classification model is built on the basis of a set
        of documents which have been previously classified into
        a number of predefined classes. How the classification

of these documents was performed is not part of the
claimed invention; they may have been classified
manually or by some (other) computerised classification
method.

3.3   The next step in building the model is the calculation
        of hyperplanes that separate the vector space into
        "subspaces" such that "each subspace comprises one or
        more documents as represented by their corresponding
        vectors" and "each subspace corresponds to a respective
        class". In other words, hyperplanes are calculated that
        bound a number of subspaces in such a way that each
        "cloud" of vectors corresponding to a particular class
        of documents lies within one subspace. These subspaces
        form the classification model. A simplified two-
        dimensional example of a separation into subspaces is
        given in Figure 3 of the application.

3.4   Once the classification model has been built, an
        unclassified document is classified by representing it
        as a vector in the same vector space and determining
        the subspace to which the vector belongs. The document
        is then classified into the class corresponding to this
        subspace.

3.5   The invention according to claim 1 further calculates a
        "maximum margin", which is a margin surrounding the
        calculated hyperplanes that does not contain any of the
        vectors corresponding to the previously classified
        documents. Upon classifying a document, a "confidence
        level" is assigned to the document based on the
        distance of its vector from the calculated hyperplanes
        relative to the corresponding margins.

4.   *Sufficiency of disclosure*

4.1   In the communication accompanying the summons, the
        Board noted that it was willing to accept that the
        skilled person was able to implement the calculation of
        a suitable set of hyperplanes *if such a set exists*, but
        that in general this appeared not to be the case.
        Indeed, if in the example of Figure 3 the vector
        corresponding to one of the previously classified
        documents in predefined class I (represented by
        squares) had happened to be located in the middle of
        the vectors corresponding to the previously classified
        documents in class IV, a set of hyperplanes with the
        properties required by the claim would not exist.

4.2   This problem of non-linearly separable clouds is in
        fact foreseen in the description on page 14, last two
        paragraphs. This passage discloses that "preferably" a
        further preprocessing is performed which has two tasks,
        the second task involving removing or merging
        underpopulated classes and splitting up overpopulated
        classes into subclasses. The method of choice for
        performing the class split is referred to as
        "clustering (unsupervised learning)". This is said to
        ensure that the generated subclasses will be pairwise
        linearly separable.

4.3   A further reference to this problem can be found on
        page 21, first paragraph, which states that for
        "overlapping classes (classes which are not pairwise
        linearly separable)" a "network growth algorithm" or
        "kernel methods (support vector machines)" can be used.

4.4   The application does not explain any of these
        techniques in detail, and claim 1 does not specify any
        measure being taken to ensure linear separability. It

may therefore be questioned whether the application is sufficiently disclosed over the whole scope claimed. However, the Board considers that this issue does not prevent it from examining for the presence of an inventive step. Given the outcome of this examination (see points 5.1 to 5.8 below), the question of sufficiency of disclosure need not be answered.

5.   *Inventive step*

5.1   Claim 1 defines a method for classifying text documents essentially in terms of an abstract mathematical algorithm. Claim 1 does specify that the algorithm is to be executed by a computer, but only by referring to steps of the method as being "computerized" and by referring to text documents as being "digitally represented in a computer".

5.2   A mathematical algorithm contributes to the technical character of a computer-implemented method only in so far as it serves a technical purpose (see decision T 1784/06 of 21 September 2012, reasons 3.1.1). In the present case, the algorithm serves the general purpose of classifying text documents.

Classification of text documents is certainly useful, as it may help to locate text documents with a relevant cognitive content, but in the Board's view it does not qualify as a technical purpose. Whether two text documents in respect of their textual content belong to the same "class" of documents is not a technical issue. The Board notes that the same position was taken in decision T 1316/09 of 18 December 2012, reasons 2, which held that methods of text classification *per se* did not produce a relevant technical effect or provide a technical solution to any technical problem.

5.3   In the statement of grounds of appeal, the appellant
        stressed that the claimed invention could not be seen
        as the straightforward implementation of something
        which had been done manually before. When manually
        classifying a text document, a human being would read
        it through and assign a particular class to it on the
        basis of his understanding of the document. As was
        known from the domain of cognitive psychology, he would
        not consider all of the words in the document; words
        near its beginning would often already provide a clear
        indication of its semantic topic. The claimed automatic
        classification method on the other hand involved
        precise computation steps which no human being would
        ever perform when classifying documents.

        The appellant also submitted that the claimed
        computerised method was highly efficient, in particular
        in comparison to classification methods disclosed in
        documents cited in the international search report.

5.4   The Board agrees that a human being would not apply the
        claimed classification method to perform the task of
        classifying text documents. The Board further accepts
        that the proposed computerised method may be faster
        than classification methods known from the prior art.
        However, the determination of the claim features which
        contribute to the technical character of the invention
        is made, at least in principle, without reference to
        the prior art (cf. T 154/04, OJ EPO 2008, 46, reasons
        5, under (E) and (F)). It follows that a comparison
        with what a human being would do or with what is known
        from the prior art is not a suitable basis for
        distinguishing between technical and non-technical
        steps (see also decision T 1954/08 of 6 March 2013,
        reasons 6.2).

5.5   Nevertheless, not all efficiency aspects of an
        algorithm are by definition without relevance for the
        question of whether the algorithm provides a technical
        contribution. If an algorithm is particularly suitable
        for being performed on a computer in that its design
        was motivated by technical considerations of the
        internal functioning of the computer, it may arguably
        be considered to provide a technical contribution to
        the invention (cf. T 258/03, OJ EPO 2004, 575, reasons
        5.8). However, such technical considerations must go
        beyond merely finding a computer algorithm to carry out
        some procedure (see G 3/08, OJ EPO 2011, 10, reasons
        13.5 and 13.5.1).

        In the present case the Board considers that no such
        technical considerations are present. The algorithm
        underlying the method of claim 1 does not go beyond a
        particular mathematical formulation of the task of
        classifying documents. The aim of this formulation is
        clearly to enable a computer to carry out this task,
        but no further consideration of the internal
        functioning of a computer can be recognised.

5.6   The appellant further argued that the claimed method
        provided more reliable and objective results than
        manual classification, since it was independent of the
        human subjective understanding of the content of the
        documents.

        The Board does not contest that the claimed
        classification method may provide reliable and
        objective results, but this is an inherent property of
        deterministic algorithms. The mere fact that an
        algorithm leads to reproducible results does not imply
        that it makes a technical contribution.

5.7   Since the mathematical algorithm does not contribute to
        the technical character of the claimed method, an
        inventive step can be present only in its technical
        implementation. The only implementation features
        specified in the claim are references to the method
        being "computerized" and the text documents being
        "digitally represented in a computer". The skilled
        person, when given the task of implementing the
        algorithm, would certainly have chosen to represent
        text documents "digitally in a computer".

        The Board further considers that the skilled person,
        using only his common general knowledge, would have had
        no difficulty in implementing on a computer the various
        steps of claim 1. The appellant never argued otherwise,
        which is consistent with the fact that the description
        of the present application does not provide any
        technical implementation details at all. Although the
        Board, as explained above in points 4.1 to 4.4, does
        have some doubts as regards the general feasibility of
        calculating a suitable set of hyperplanes, these
        concerns essentially relate to the definition of the
        algorithm and not to its implementation.

5.8   The technical implementation of the mathematical
        algorithm being obvious, the conclusion is that the
        method of claim 1 lacks an inventive step within the
        meaning of Articles 52(1) and 56 EPC over a notorious
        general-purpose computer.

6.   Since the sole claim request on file is not 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