

Internal distribution code:

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

**Datasheet for the decision
of 26 June 2008**

Case Number: T 1569/05 - 3.5.01
Application Number: 97305783.9
Publication Number: 0822506
IPC: G06F 17/30
Language of the proceedings: EN

Title of invention:

Method and apparatus for retrieving data

Applicants:

CANON KABUSHIKI KAISHA
Kitagawa, Takashi
Kiyoki, Yasushi

Opponent:

-

Headword:

Method for retrieving data/CANON

Relevant legal provisions:

EPC Art. 52(1), 123(2)

Relevant legal provisions (EPC 1973):

EPC Art. 56, 84
EPC R. 29(1)

Keyword:

"Inventive step (no)"

Decisions cited:

T 0208/84, T 0052/85, T 1351/04

Catchword:

See point 3 of the Reasons.



Case Number: T 1569/05 - 3.5.01

D E C I S I O N
of the Technical Board of Appeal 3.5.01
of 26 June 2008

Appellants:

CANON KABUSHIKI KAISHA
30-2, 3-chome, Shimomaruko,
Ohta-ku
Tokyo (JP)

Kitagawa, Takashi
207-202, Namiki 2-chome Tsukuba-shi,
Ibaraki-ken (JP)

Kiyoki, Yasushi
1148-24, Nagakuni Tsuchiura-shi,
Ibaraki-ken (JP)

Representative:

TBK-Patent
Bavariaring 4-6
80336 München (DE)

Decision under appeal:

Decision of the Examining Division of the
European Patent Office posted 4 August 2005
refusing European application No. 97305783.9
pursuant to Article 97(1) EPC 1973.

Composition of the Board:

Chairman: S. Steinbrener
Members: S. Wibergh
P. Schmitz

Summary of Facts and Submissions

- I. This appeal is against the decision of the examining division to refuse European patent application No. 97305783.9.
- II. According to the decision appealed, the invention did not involve an inventive step having regard to document D3: Y. Kiyoki et al., "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning", SIGMOD RECORD, Vol. 23, No. 4, December 1994, pages 34-41.
- III. With the statement setting out the grounds of appeal, dated 14 December 2005, the appellants requested that the decision be set aside and a patent be granted based on claims 1-26 of the main request, auxiliary request 1 or auxiliary request 2 filed with the same letter.
- IV. In a communication, the Board stated that it was doubtful if the method solved a technical problem. Search methods involving control signals, eg in the form of trees, and key word comparisons in order to direct the processor to the desired data item would appear to be fundamentally patentable. The situation was however different if the lexical meaning of the search items was decisive. An important question was whether the use of a model to search a database rendered the model a technical feature in the sense of Rule 29(1) EPC 1973, or whether the effect of, for example, retrieving images (embodiment 3) was merely the physical manifestation of the results delivered by the mathematical model, similar to the effects of

searching a database manually. A mathematical model could only be taken into account for patentability if it clearly related to a technical application.

The Board was also of the opinion that the definition of the "principal-axis index set" in claims 1 and 13 of all requests was obscure and not fully supported by the original description.

V. By letter dated 23 May 2008, the appellants filed an amended set of claims according to a new main request. The previous requests were maintained as auxiliary requests 1-3.

VI. Claim 13 or the main request reads:

"A computer-implemented semantic data processing method performed by a computer to search a database for data, said method comprising:

a first inputting step (S1801) of inputting a keyword;
a space generation word transforming step (S401, S402) of transforming a predetermined space generation word group into a space generation vector group by transforming each space generation word of the space generation word group into a space generation vector which has elements corresponding to a predetermined characteristic word group to represent a meaning of the space generation word;
a semantic space generation step (S203, S402, -S407 /sic/) of generating a semantic space on the basis of the space generation vector group obtained in said space generation word transforming step;
a second inputting step (S1301) of inputting a context word group;

a first transforming step (S1802-S1804) of transforming the keyword into a keyword vector in the semantic space, the keyword vector corresponding to a combination of words which are used to describe a meaning of said keyword in a dictionary where meanings of words are described by predetermined basic words;

a second transforming step (S1302) of transforming each context word in the context word group into a vector in a context word vector group in the semantic space, the vector corresponding to a combination of words which are used to describe a meaning of said context word in said dictionary;

a third inputting step (S1601) of inputting a comparison-subjected vector group in the semantic space, each vector in the comparison-subjected vector group corresponds to respective data in the database (205, 305);

a semantic center calculating step (S1306) of calculating a semantic center of the context word vector group by performing a logical operation on all vectors of the context word vector group and dividing the results of the logical operation by a norm thereof;

a projector generating step (S1307, S1308) of generating a projector for projecting a vector in the semantic space into a substance */sic, should be* subspace/ of the semantic space corresponding to the context word group, on the basis of the semantic center;

a projecting step (S1603, S1804) of projecting the keyword vector and the comparison-subjected vector group in the substance */sic, should be* subspace/ by utilizing the projector;

a calculating step (S1805) of calculating a correlation amount between each word of a comparison-subjected word group and the keyword; and

a selecting step (S2309, S2809) of selecting at least one vector from the comparison-subjected vector group on the basis of the correlation amount; and
a retrieving step (S2310, S2810) of retrieving data from the database (205, 305) based on the selected vector and outputting the retrieved data as a search result,
characterized in that,
in said second inputting step, the comparison-subjected vector group is input by transforming each comparison-subjected word in a comparison-subjected word group into a vector in the comparison-subjected vector group in the semantic space, the vector corresponding to a combination of words which are used to describe a meaning of said comparison-subjected word in said dictionary;
in said semantic space generation step, a principal-axis index set is generated (S407) by calculating a sum vector of the space generation words /sic/ vector group and selecting an axis of the sum vector as the principal-axis index set if an absolute value of corresponding element satisfies a condition for a ratio to an absolute value of a succeeding element in descending order of the absolute values;
in said projector generating step, the projector is generated (S1307, S1308) so as to project the vector in the subspace consisting of axes that correspond to elements of the semantic center, the absolute values of which are larger than a predetermined value, and that do not belong to the principal-axis index set, and
in said retrieving step, data associated with a word corresponding to the selected vector in the database are retrieved".

Claim 1 is directed to a corresponding "semantic data processing apparatus for searching database for data".

- VII. According to auxiliary requests 1-3, claim 13 differs from the main request by minor amendments. All the auxiliary versions of the claim include the following formulation in respect of the generation of the principal-axis set:
- "selecting a plurality of axes in descending order of the absolute values of a sum of the space generation vector group for respective axes".
- VIII. Oral proceedings were held on 26 June 2008. The appellants requested that the decision under appeal be set aside and that a patent be granted on the basis of the main request filed with letter of 23 May 2008 or on the basis of auxiliary requests 1 to 3 filed with letter of 14 December 2005 as then main request and auxiliary requests 1 and 2.
- IX. At the end of the oral proceedings the Board announced its decision.

Reasons for the Decision

The main request

1. The invention

The invention is a data processing method (claim 13) and apparatus (claim 1) for searching a database. The stored data could be of any kind but for illustration it is here assumed they represent images in accordance with the

third embodiment of the invention. Each image is described by a number of words ("comparison-subjected word group") representing its contents. A user searching for an image inputs a keyword as well as a number of "context words" intended to define the appropriate semantic context. The keyword, the context words and the comparison-subjected words are transformed to vectors in what is referred to as "semantic space". This space has been created using eigenvalue decomposition of "space generation words", taken for example from a dictionary. The context vectors form a "semantic center", which is a subspace of semantic space corresponding to the given context. The semantic center does not include the "principal axes" of semantic space, ie the axes corresponding to the most frequent meanings of the space generation words. The keyword vector and the comparison-subjected vector group are projected onto the semantic center and the distances ("correlation amounts") between the keyword vector and the comparison-subjected vectors are calculated. The closest comparison-subjected vector is identified and the corresponding image retrieved from the database (see also p.4, 1.44 to p.11, 1.9 of the A-publication).

2. Technicality

Claim 13 is directed to a "computer-implemented method... performed by a computer". A computer being a technical means, the subject-matter of claim 13 is an invention within the meaning of Article 52(1) EPC.

3. Inventive step

3.1 The appellants accept that D3, presenting earlier work by the inventors, discloses all major features of the preamble of claim 13 and also the following characterising features:

- the comparison-subjected vector group is input by transforming each comparison-subjected word in a comparison-subjected word group (item set W; p.39, left column) into a vector in the comparison-subjected vector group in the semantic space, the vector corresponding to a combination of words which are used to describe a meaning of said comparison-subjected word in said dictionary;
- the projector is generated so as to project the vector into the subspace consisting of axes that correspond to elements of the semantic center, the absolute values of which are larger than a predetermined value (page 38, section 3.3, last equation), and
- data associated with a word corresponding to the selected vector in the database are retrieved (p.39, last paragraph of section 4).

3.2 The subject-matter of claim 13 thus differs from the teaching of D3 in that:

- a principal-axis index set is generated by calculating a sum vector of the space generation vector group and selecting an axis of the sum vector as the principal-axis index set if an absolute value of the corresponding element satisfies a condition for a ratio

to an absolute value of a succeeding element in descending order of the absolute values, and
- the subspace into which the projector projects a vector contains no axes belonging to the principal-axis index set.

As to the arguably obscure formulation of the first distinguishing feature, the appellants submit that its true meaning is clear from figures 10 and 11 of the application. For the purposes of the present decision the feature is therefore taken to mean that axes corresponding to sufficiently large components of the sum vector are selected as principal axes.

Hence, in essence the claimed data processing method differs from the prior art by a modification of the mathematical model of meaning used for data retrieval. Put simply, common elements of meaning, having no distinguishing power, are determined, and the corresponding axes are excluded from the subspace ("semantic center") where the correlations between the keyword and the image descriptions ("comparison-subjected word group") are evaluated.

3.3 Also the examining division found that the above two features (as they were then formulated) distinguished the invention from D3 (cf the decision under appeal, point 1.1). In the division's opinion, the features merely caused a further restriction of the subspace to be searched (cf the decision under appeal, point 1.2). This was a technically non-functional modification of the known "mathematical model of meaning", relating to the field of linguistics. The invention thus did not involve an inventive step.

- 3.4 The appellants regard the invention as belonging to the technical field of utilizing a natural language as a search input. In their view, the invention makes a technical contribution in a field not excluded from patentability, as required by decision T 208/84 - Computer-related invention/VICOM (OJ EPO 1987,14), and by eliminating the negative influence of elements shared by many words it renders the retrieval both more efficient and more correct than the model in D3.
- 3.5 In the Board's view, neither the mathematical model of meaning according to D3 nor the modified model according to the invention are within the technical area, since only the meaning of the words determines how they are represented, stored and selected, and since mathematical algorithms completely define the processing. In this respect the present invention is similar to the case T 52/85 - Listing of semantically related expressions/IBM (not published in OJ EPO), where the deciding board held that automatically generating a list of expressions semantically related to an input linguistic expression is basically not of a technical nature but a matter of the meaning of those expressions, ie of their abstract linguistic information content.
- 3.6 A technical aspect can therefore at most be seen in the application of these models for retrieving data in a computer database, such retrieval being normally considered to have technical character.
- 3.7 In the present case the retrieving step produces a different result than the prior art for the sole reason

that the semantic subspace used for the retrieval has been scaled down. Hence, the only principles that have been modified concern the search for the image description closest in meaning to the desired description (keyword). They do not concern the search performed within the database to retrieve the image corresponding to the input data.

This distinguishes the present invention from the subject-matter considered in decision T 1351/04 - File search method/FUJITSU (not published in OJ EPO). In that case the board saw a technical effect in "the control of the computer along the path leading to the desired data" (point 7.2). In the present case, however, the search is not primarily for a certain data location but for certain words having a given lexical meaning. On the basis of these words the computer finds the associated images, but how this is done is not part of the invention.

The present situation is also different from that in decision T 208/84, referred to by the appellants. This decision states that a technical process "is carried out on a physical entity (which may be a material object but equally an image stored as an electric signal) by some technical means implementing the method and provides as its result a certain change in that entity" (cf reasons, point 5). In the present case there is however no change in a physical entity. The stored data are not modified but merely retrieved. Hence, decision T 208/84 is not applicable.

- 3.8 As has been pointed out above, the basic principles of the mathematical model of meaning and its application for data retrieval purposes are known from document D3.

The claimed modification of this model relates to a reduction of the initial data set and may therefore allow faster or better correlation results of meanings. Having regard to semantics, it goes without saying that features lacking any distinctiveness are not helpful for a comparison of meanings. From the linguistic point of view, it would therefore be natural to do without them since such features only form useless ballast in the meaning analysis. Therefore, even without an express hint in D3, an expert team, possibly including a linguist and a mathematician, would be aware that very frequent characteristics are not distinguishing and thus should be eliminated. The expert team would understand from the definition of the known semantic space model that the mathematical equivalent of removing frequent characteristics is to omit the axes along which most data are concentrated, ie the "principal axes".

To use such a modified model for data retrieval is obvious in the light of D3. Search efficiency is a standard problem in data retrieval applications and any modification leading to faster and arguably better search results would be clearly desirable. Hence, even if the modification was prompted by its use in data retrieval, the Board would not consider it inventive as the removal of useless data must necessarily improve search efficiency. The implementation of such a modification in the method according to D3 is also considered straightforward under the aspect of programming.

It follows that the subject-matter of claim 13 does not involve an inventive step (Article 56 EPC 1973).

The auxiliary requests

4. Independent claim 13 of auxiliary request 1 is not clear (Article 84 EPC 1973) and, as far as it can be understood, contains subject-matter extending beyond the content of the application as filed (Article 123(2) EPC). The claim states that "a principal-axis index set is generated (S407) by selecting a plurality of axes in descending order of the absolute values of a sum of the space generation vector group for respective axes". A skilled reader would probably interpret this passage in the way that the principal-axis set will always contain at least two axes. According to fig.10, however, it appears that the set is sometimes empty ("YES" in box S1008) or contains a single axis. A method always generating at least two principal axes has thus not been disclosed, nor have the appellants alleged that it has.

The same objections apply to claim 13 of auxiliary requests 2 and 3.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chairman:

T. Buschek

S. Steinbrener