

Internal distribution code:

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

**Datasheet for the decision
of 10 February 2021**

Case Number: T 2587/18 - 3.5.07

Application Number: 14177795.3

Publication Number: 2838037

IPC: G06F17/30

Language of the proceedings: EN

Title of invention:

Information processing system, information processing method,
and information processing program

Applicant:

Fujitsu Limited

Headword:

Search string in compressed data/FUJITSU

Relevant legal provisions:

EPC Art. 56

RPBA 2020 Art. 13(2)

Keyword:

Inventive step - main request and first and second auxiliary
requests (no)
Amendment after summons - third auxiliary request -
exceptional circumstances (no)



Beschwerdekammern
Boards of Appeal
Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0
Fax +49 (0)89 2399-4465

Case Number: T 2587/18 - 3.5.07

D E C I S I O N
of Technical Board of Appeal 3.5.07
of 10 February 2021

Appellant: FUJITSU LIMITED
(Applicant) 1-1, Kamikodanaka 4-chome,
Nakahara-ku
Kawasaki-shi,
Kanagawa 211-8588 (JP)

Representative: Hoffmann Eitle
Patent- und Rechtsanwälte PartmbB
Arabellastraße 30
81925 München (DE)

Decision under appeal: **Decision of the Examining Division of the
European Patent Office posted on 23 May 2018
refusing European patent application
No. 14177795.3 pursuant to Article 97(2) EPC**

Composition of the Board:

Chair M. Jaedicke
Members: P. San-Bento Furtado
C. Almborg

Summary of Facts and Submissions

- I. The appeal lies from the decision of the examining division to refuse European patent application No. 14177795.3. The appealed decision cited, *inter alia*, the following document:
D3: US 2004/0225497 A1, published on 11 November 2004.

The examining division decided that claim 1 of each of a main request and first and second auxiliary requests was not inventive over the disclosure of document D3.
- II. In the statement of grounds of appeal, the appellant requested that the appealed decision be set aside and that a patent be granted on the basis of one of the main request and first and second auxiliary requests considered in the appealed decision.
- III. In a communication accompanying a summons to oral proceedings, the board expressed its preliminary opinion that the subject-matter of claim 1 of each of the requests was not inventive over the disclosure of document D3 in combination with the common general knowledge of the skilled person.
- IV. In a letter of reply, the appellant provided further arguments in favour of inventive step of the claimed invention.
- V. Oral proceedings were held by videoconference as scheduled, during which the appellant submitted a third auxiliary request. At the end of the oral proceedings, the Chair announced the Board's decision.

- VI. The appellant's final requests were that the appealed decision be set aside and that a patent be granted on the basis of one of the main request and the first to third auxiliary requests.
- VII. Claim 1 of the main request reads as follows:
" An information processing system (100) characterized by comprising:
 a creating unit (602) configured to, when object data is compressed for each word in units of records, create count data that indicates for each record of the object data, an appearance count of each word, the count data being added to the object data that has been compressed, and
 an identifying unit (603) configured to identify, based on the count data, a second character string that appears a predetermined number of times in a record in which a first character string appears, as corresponding character string of the first character string, the first character string being defined as a search condition for the object data."
- VIII. Claim 1 of the first auxiliary request reads as follows:
"An information processing method comprising:
 a step of creating count map data from object data when the object data is compressed, the object data including a plurality of record units, the object data being compressed by units of words, the count map data being created for each of the plurality of record units, respectively, and indicating appearance counts of words included in the each of the plurality of record units;
 a step of adding the count map data to compressed object data that has been compressed from the object data;

a step of specifying a first character string from a search condition for the object data; and

a step of identifying a second character string that has a co-occurrence relationship with the first character string in the object data utilizing the count map data."

- IX. Claim 1 of the second auxiliary request differs from claim 1 of the main request in that the following text is added at the end of the claim:

",". [sic] wherein

the creating unit (602) is further configured to create and add to the compressed object data, classification data that indicates a type of each word, and

the identifying unit (603) is further configured to identify, based on the count data and the classification data, the second character string that corresponds to a classification and the first character string defined as search conditions for the object data."

- X. Claim 1 of the third auxiliary reads as follows:

"An information processing system (100) characterized by comprising:

a creating unit (602) configured to, when object data (102) is compressed for each word in units of records, create count data (105) that indicates for each record of the object data (102), an appearance count of each word, the count data being added to the object data (102) that has been compressed, and

an identifying unit (603) configured to identify, based on the appearance count of each word in a record in which the appearance count of a first character string (101) is one or more according to the count data (105) , a second character string (103) that appears a

predetermined number of times in a record in which the first character string (101) appears, as corresponding character string of the first character string (101), the first character string (101) being defined as a search condition for the object data (102);

a calculating unit (605) configured to calculate a significance of the identified second character string (103) based on the count data (105) and the number of records in the object data (102); and

an output unit (606) configured to output the identified second character string (103) and the respective significance."

- XI. The appellant's arguments, where relevant to this decision, are addressed in detail below.

Reasons for the Decision

Invention

1. The invention concerns searching of compressed object data for a second character string (e.g. a word) corresponding to a first character string given as input. The object data may be a document, or one or more groups of documents and is divided into multiple records. A record may be a section in a document or a document in a group of documents (see application as originally filed, page 3, last full paragraph, to page 4, first full paragraph).

The search method according to the invention identifies a second character string which appears at least a predetermined number of times together with the first character string in a record of the compressed object data without decompressing the data. This is achieved by providing, in addition to the compressed data, a

"count map" ("count data" in the claims) that indicates, for each record of the object data, the number of times each word appears in the record (page 4, second full paragraph, to page 5, second full paragraph, Figure 1).

Main request

2. *Inventive step - claim 1*

2.1 Document D3 discloses a system for storing and retrieving text, in which the text is compressed. The first step in creating the compressed file is to break up the text into "items". The text is then compressed by assigning a 16-bit word identification number (WID) to each unique word in the text, creating a tokenised text file (TTF), and building a word table (equivalent to a token dictionary). Each item can be decompressed separately. Further compression steps may then be performed (paragraphs [0009] to [0011], [0023] and [0024], Figures 1 and 2). In order to support fast searches, an index with a given text resolution for each unique word is created and added as the second column element in the alphabetised word table. The index for a word indicates in which "index spans" of the text the word is present, where an index span may be set as a single item or a number of items. For example, if a word occurs in the first and second index spans of a text consisting of two index spans, then the index is "11", if the word occurs only in the second index span, the index is "01". The use of this WID index makes it possible to search for a keyword without having to decompress the file (paragraphs [0011] and [0032], Figure 6).

2.2 The board agrees in its essence with the examining division's assessment of inventive step. Document D3 discloses a system comprising a creating unit configured to, when object data (text in D3) is compressed for each word in units of records (where an index span corresponding to one item in D3 may be seen as a record), create index data that indicates for each record of the object data the appearance of each word, the index data being added to the compressed object data. Document D3 also discloses using the index data for searching the object data for the records including a first character string, the first character string being defined as a search condition (paragraph [0038], Figure 8). Document D3 thus implicitly discloses an identifying unit configured for that search.

2.3 The subject-matter of claim 1 therefore differs from the system of D3 in that

- (i) the count data (in place of the index data of D3) indicates an appearance count instead of an appearance;
- (ii) the identifying unit identifies, based on the count data, a second character string that appears a predetermined number of times in a record in which the first character string appears.

In the statement of grounds of appeal, the appellant argued that the distinguishing features had the technical effect of saving time and processing resources and solved the technical problem of "how to determine, from a data object, a second string appearing a predetermined number of times in a record corresponding to a search condition without decompressing the data object".

The board notes that the claim does not disclose how the count data is used for identifying the second character string. However, at least for the sake of argument, the board interprets the distinguishing features as implying that the search defined in (ii) is performed in an efficient manner without decompressing the data object.

In that case, the distinguishing features solve the technical problem of modifying the system of D3 to efficiently identify a second character string that appears a predetermined number of times in a record in which the first character string appears.

- 2.4 The appellant argued that nothing in D3 guided the skilled person towards modifying the system disclosed in D3 in the direction of features (i) and (ii). In document D3, counting was used only for sorting the frequencies, but counting information was not saved and was therefore lost. The appellant further argued that the skilled person would have regarded text compression and text utilisation as different problems and would thus have concentrated on either of the problems. Document D3 aimed at achieving high compression rates, as was clear for instance from paragraphs [0007], [0027], [0029], [0030] and [0034], and the index of Figure 6 was compressed. It was clear from paragraph [0034] that the index of D3 was sparse and consisted of "1"s and "0"s, and that the system of D3 relied on efficiently compressing the index using run-length-encoding (RLE). The skilled person would not have used counts of occurrences instead of binary values in the index because that would be contrary to the aim of efficiently compressing the index.

2.5 The Board does not find the appellant's arguments convincing. At the priority date of the present application, the skilled person designing a compression scheme for text data would already have taken into account how the text data is going to be used later. This is illustrated in document D3, which is about searching compressed texts for keywords without having to scan the (compressed) text data. The index is created for that purpose (see paragraphs [0011], [0012], [0032], [0037] and [0038]).

When confronted with the above mentioned technical problem, the skilled person would immediately have recognised that it would be very inefficient to use the index of the system of D3 to search for second character strings that appear a predetermined number of times in a record (index span in D3) in which the first character string appears. The index of D3 would have had to be searched to identify all the records containing both the first and second character strings and then each of the identified records would have had to be decompressed to find out how often the second character string appeared in each of the identified records.

The skilled person would have been aware, as it was commonly known and anyway taught in D3, that decompression should be avoided for performing the search efficiently. The skilled person would have considered achieving that by a modification of the index of D3 to support the new search without decompressing the records, especially since D3 discloses using the index for searching words without decompression. Since the search mentioned in the problem faced by the skilled person requires the information regarding the number of appearances of the

second character string in a record, it would have been obvious to modify the index of D3 to indicate the number of appearances of a character string in a record instead of merely indicating the appearance or non-appearance of a character string in a record. Contrary to the appellant's arguments, storing the count data would have been an obvious option.

The board is not persuaded by the appellant's argument that the skilled person would not have modified the index in that way because the index is efficiently compressed using RLE in the system of D3. The compression of the index is disclosed as an optional feature in document D3. Moreover, the skilled person would have known that an index with count numbers can also be efficiently compressed with RLE and would have considered avoiding decompression of the text when performing the particular search more important than efficiently compressing the index.

- 2.6 After discussing inventive step of the main request and first and second auxiliary requests at the oral proceedings, the appellant further argued that the invention also solved an alternative objective technical problem related to minimising the number of times parsing is performed.

Before the oral proceedings, the appellant had not contested the technical problem formulated by the board, which is similar to that formulated by the appellant in the statement of grounds of appeal. In the oral proceedings, the appellant confirmed that it did not disagree with that technical problem formulated by the board, but argued that the claimed invention also solved the alternative technical problem. According to the appellant, the claimed invention was inventive because document D3 required much more parsing than the

claimed invention, and there was no pointer in the prior art to the distinguishing features of claim 1.

Since in its assessment the board concludes that the invention lacks inventive step over document D3 on the basis of a technical problem solved over D3 which was not contested by the appellant, this additional argument based on a second technical problem is not relevant. As clearly established in the case law, if the invention is obvious to the skilled person in respect of *at least one* of different routes starting from different documents, then an inventive step is lacking (Case Law of the Boards of Appeal, 9th edition, July 2019, I.D.3.1). The same holds true for different routes based on different problems starting from the same prior art.

- 2.7 Therefore, the subject-matter of claim 1 of the main request is not inventive (Article 56 EPC).

First auxiliary request

3. Claim 1 of the first auxiliary request specifies a method comprising the steps of creating count-map data, adding the count-map data to compressed object data, specifying a first character string as search condition, and identifying a second character string that has a co-occurrence relationship with the first character string in the object data utilizing the count-map data.

Those features of claim 1 correspond to features defined in claim 1 of the main request, except that the second character string is identified on the basis of a mere co-occurrence relationship with the first character string, whereas in claim 1 of the main request the condition is that it appears a

predetermined number of times in a record in which the first character string appears.

4. *Inventive step - claim 1*

4.1 Contrary to the appellant's arguments, document D3 also discloses, apart from feature (i), steps corresponding to the steps of claim 1 of the first auxiliary request of creating count-map data (creating the index in D3), and adding the count-map data (the index) to the compressed object. The subject-matter of claim 1 of the first auxiliary request differs from the method of D3 in that it includes features (i) and (iii) the step of identifying a second character string that has a co-occurrence relationship with the first character string in the object data utilizing the count-map data.

4.2 The appellant argued that the method of claim 1 of the first auxiliary request solved the objective technical problem of determining, from a data object including a plurality of record units, a second character string that has a co-occurrence relationship with the first character string without decompressing the data object. The claimed method was inventive for the same reasons as given for the main request.

4.3 The board notes that the index of document D3 can be used to perform step (iii) without decompressing the records. Therefore, the count-map data of feature (i) no longer has an advantage over the index of D3 in the context of performing a co-occurrence search as described in feature (iii). Feature (i) can therefore be treated independently of feature (iii) in the inventive-step assessment. In the Board's opinion, feature (i) has the advantage of efficiently supporting searches on the basis of the number of occurrences of a

string in a record. As explained for the main request, it would have been obvious for the skilled person seeking to achieve that advantage to modify the index of D3 to indicate the number of appearances of a character string in a record instead of merely indicating the appearance or non-appearance of a character string in a record.

Feature (iii) solves the problem of efficiently identifying a second character string that appears in a record in which the first character string appears. It would have been a matter of ordinary programming skills, especially when starting from D3, to use count data (or the index data of D3) for implementing such a search without decompressing data.

- 4.4 In sum, claim 1 of the first auxiliary request is not inventive (Article 56 EPC).

Second auxiliary request

5. Claim 1 of the second auxiliary request differs from claim 1 of the main request in that it specifies that
- (iv) the creating unit creates and adds classification data to the compressed object data which indicates a type of each word and
 - (v) the identified second character string corresponds to a classification also defined as search condition.

6. *Inventive step - claim 1*

- 6.1 In accordance with the description and drawings of the application, the classification data may give indication, for instance, of whether the string is an adjective (see e.g. Figure 4).

The appellant argued that the additional features of the second auxiliary request were technical and the distinguishing features solved the technical problem of determining from a data object a second string appearing a predetermined number of times and having a determined classification without decompressing the data object.

In the board's opinion, the classification data is, in the light of the description, non-technical linguistic information and there is no technical reason in the present case for retrieving data meeting a criterion on the type of word. Therefore, it is legitimate to include in the technical problem the requirement of retrieving data on the basis of the classification (as the appellant did in its formulation of the technical problem).

The distinguishing features solve the technical problem of modifying the system of D3 to efficiently identify a second character string that has a determined classification and appears a predetermined number of times in a record in which the first character string appears.

As explained for the main request, it would have been obvious for the skilled person, especially in view of the teaching of document D3, to implement the new search efficiently by avoiding decompression. In order to achieve that, the skilled person would immediately have considered adding the classification data to the index of D3, especially since document D3 already discloses using the index data for searching without decompression. Therefore, it would not have been inventive to add features (iv) and (v) to the system of document D3. As explained with regard to the main request, the skilled person confronted with the above

mentioned problem would also have considered adding features (i) and (ii) to the system of D3.

6.2 Therefore, the second auxiliary request does not fulfil the requirements of Article 56 EPC.

Third auxiliary request

7. Claim 1 of the third auxiliary request differs from claim 1 of the main request in that reference signs were added, it is specified that the identifying unit identifies the second character string "based on the appearance count of each word in a record in which the appearance count of a first character string (101) is one or more according to the count data", and in that it includes

- a calculating unit configured to calculate a significance of the identified second character string based on the count data and the number of records in the object data; and
- an output unit configured to output the identified second character string and the respective significance.

8. *Admission of the request into the proceedings*

8.1 The third auxiliary request was filed during the oral proceedings and hence in a late phase of the already advanced stage of the appeal proceedings after notification of a summons to oral proceedings as referred to in Article 13(2) RPBA 2020.

The appellant argued that the late submission of this auxiliary request was justified and that the request should be admitted because it had come out for the first time at the oral proceedings that the skilled person would have considered modifying the system of D3

in order not to compress the index. Furthermore, the claim overcame all the outstanding objections.

These arguments are not convincing. Document D3 was the starting point of the inventive step assessment in the appealed decision. In point 4.4 of its preliminary opinion, the board had already expressed how the skilled person would have modified D3, which was then repeated during the oral proceedings. The argument based on index compression in the system of D3 was brought up by the appellant for the first time at the oral proceedings. Therefore, the board does not recognise the presence of any exceptional circumstances which would justify admitting the request. In addition, the third auxiliary request does not *prima facie* overcome the objection of inventive step.

- 8.2 In view of this, in accordance with Articles 13(1) and 13(2) RPBA 2020, the board does not admit the third auxiliary request into the appeal proceedings.

Concluding remark

9. Since the main request and first and second auxiliary requests are not allowable and the third auxiliary request is not admitted into the proceedings, the appeal is to be dismissed.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chair:



S. Lichtenvort

M. Jaedicke

Decision electronically authenticated