

Internal distribution code:

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

**Datasheet for the decision
of 15 March 2013**

Case Number: T 1465/11 - 3.5.01

Application Number: 04253257.2

Publication Number: 1508864

IPC: G06F17/30

Language of the proceedings: EN

Title of invention:

Apparatus and method for searching data of structured document

Applicant:

FUJITSU LIMITED

Headword:

Searching with automata/FUJITSU

Relevant legal provisions:

EPC 1973 Art. 56

Keyword:

Inventive step - (no)



**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 1465/11 - 3.5.01

D E C I S I O N
of Technical Board of Appeal 3.5.01
of 15 March 2013

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

Representative: Stebbing, Timothy Charles
Haseltine Lake LLP
Lincoln House, 5th Floor
300 High Holborn
London WC1V 7JH (GB)

Decision under appeal: **Decision of the Examining Division of the
European Patent Office posted on 29 December
2010 refusing European patent application No.
04253257.2 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman: S. Wibergh
Members: P. Scriven
A. Pignatelli

Summary of Facts and Submissions

- I. The appeal is against the Examining Division's decision to refuse European patent application 04253257.2 for lack of inventive step, because the invention would have been obvious in the light of document D3 (Green et al., *Processing XML streams with deterministic automata*, 9th International Conference on Database Theory, 8 - 10 January 2003, Springer-Verlag, Berlin 2003).
- II. In the statement setting out its grounds of appeal, the appellant requested that the appealed decision be set aside and that a patent be granted on the basis of a main or else one of five auxiliary requests, all filed with the statement of grounds, or that oral proceedings be held, if the Board intended to issue a decision which was not in compliance with those requests. The appellant also requested that the Board communicate its preliminary view as early as possible.
- III. The Board arranged for oral proceedings to be held. In an annex to the summons, the Board presented its preliminary opinion.
- IV. The appellant responded to the summons by filing further arguments, and three new auxiliary requests to replace the five filed with the statement of grounds.
- V. Oral proceedings were held as scheduled. During the proceedings, the appellant filed a new, fourth auxiliary request. The final requests were that the decision to refuse the application be set aside, and that a patent be granted on the basis of the main request, filed with the statement of grounds, or on the basis of one of auxiliary requests 1 to 3, filed on 31 January 2013, or on the

basis of auxiliary request 4, filed during oral proceedings before the Board.

VI. Claim 1 according to the main request reads as follows.

*A search apparatus (201) which searches data of a document structured using a tag based on a search condition, comprising:
a generation device (101 and 213) operable to analyze the search condition and to generate a tag automaton (104 and 222) including a registered character string of a tag indicating each element contained in a search path specified by the search condition, a keyword automaton (105 and 224) including a registered character string of a keyword specified by the search condition, and state management information (106 and 223) for management of a current search state using the search path, before searching a structured document;
a read device (102 and 215) operable to sequentially read document data of the structured document to be searched by a predetermined amount See §21, which has read ..., in predetermined amounts; §23; and
a search device (103 and 214) operable to search the document data of the structured document and to output a search result while switching between a tag search of detecting a tag registered in the tag automaton from among a read data string and a keyword search of detecting a keyword registered in the keyword automaton from among the read data string using the state management information, wherein the state management information*

contains information for management of correspondence between the current search state and an element in the search path, and information designating a state in which the tag search is to be switched into the keyword search, and the search device refers to the state management information when the registered tag is detected by the tag search and switches from the tag search to the keyword search when the current search state is the state in which the tag search is to be switched into the keyword search.

- VII. Claim 1 according to auxiliary request 1 reads identically, except for the following excerpts (emphasis by the Board to indicate modifications).

...

*a generation device (101 and 213) operable to analyze the search condition and to generate **according to information extracted from the search condition** a tag automaton (104 and 222) including **registered character strings of tags** indicating each element contained in a search path specified by the search condition, **the path specifying an element to be searched for**, a keyword automaton*

...

*a search device (103 and 214) operable to search the document data of the structured document and to output a search result while switching between a tag search of detecting a tag registered in the tag automaton from among a read data string **by inputting the read document data character by character to the tag automaton**, and a keyword search of detecting a*

keyword registered in the keyword automaton from among the read data string **by inputting the read document data subsequent to the switch character by character to the keyword automaton**, using the state management information

...

wherein the state management information contains information for management of correspondence between the current search state and an element in the search path, **indicates the path specifying the element to be searched for**, and **contains** information designating a state in which the tag search is to be switched into the keyword search, and the search device refers to the state management information when **each** registered tag is detected by the tag search and switches from the tag search to the keyword search when the current search state is the state in which the tag search is to be switched into the keyword search, **the state in which the tag search is to be switched into the keyword search being the state when the detected tag indicates the element to be searched for**.

VIII. Claim 1 according to the second auxiliary request adds the following text at the end.

...; and
said apparatus is operable to:
read in the tag search a character string character by character from the document data of the structured document;
employ the tag automaton to compare the read character string character by character with the registered character strings of the tags registered in the tag automaton;

check the current search state by referring to the state management information when the read character string matches a registered character string of the registered character strings of the tags;
switch from the tag search to the keyword search when the current search state is the state in which the tag search is to be switched into the keyword search;
read in the keyword search a character string subsequent to the character string matching the registered character string of the tag indicating the element to be searched for;
employ the keyword automaton to compare the read character string character by character with the registered character string of the of the keyword registered in the keyword automaton; and determine that the structured document is a candidate for a document satisfying the search condition if the read character string matches the registered character string of the keyword.

- IX. Claim 1 according to the third auxiliary request reads identically except for the suppression of the first "and" (from the additional text set out in point VIII) and the addition of the following text at the end.

...; and
the registered character strings of the tags and keyword to which the read characters string are compared were generated according to the information extracted from the search condition.

- X. Claim 1 according to the fourth auxiliary request reads identically, except that the final added clause now reads

*... and
the tags and keyword which are detected are
detected by a said comparison with a said
registered character string.*

XI. The appellant's arguments can be summarised as follows.

The invention dealt with searching for target documents. If there were many documents, the search for "a needle in a haystack" was a technical endeavour. The invention was, therefore, not simply the abstract idea of using state machines and did not simply generate information, but involved a physical apparatus, and solved technical problems such as the efficient use of memory and computing resources. The advantages of the invention went beyond simply choosing an efficient algorithm, because switching between tag and keyword searches was efficient irrespective of the complexity of the algorithm. If there were, say, a million documents, and only one had the right tag, it would not be necessary to look for the keyword in all of those that did not have the right tag. The gain in efficiency also was derived from the fact that during the tag search, the keyword automaton was not loaded into memory. That was an important distinction over the disclosure of D3, which was of a single automaton, all of which had to be in memory. Even the "lazy" DFA (deterministic finite automaton), disclosed in D3, which generated states only as needed, did not remove states from memory once they had been created. The present invention, in contrast, needed only to hold the states of the tag automaton in memory, and needed only infrequently to load the keyword automaton, if at all, if the documents that met the search criteria were rare.

The present invention was further distinguished from the disclosure of D3 in that the automaton in D3 did not scan

the documents directly, but scanned the output of a SAX parser. According to D3, all SAX events had to be detected, and then the interesting ones, from the point of view of the search, had to be found by the automaton. The present invention, however, in so far as it could be said to detect SAX events at all, only detected those that were needed for the search. In that way, even if the SAX parser in D3 were itself implemented as an automaton, the present invention achieved an advantage by not detecting and processing all SAX events, but only those that were relevant to the search. To arrive at the invention, starting from D3, the skilled person would have had to discard the SAX parser entirely, and redesign the system without one. There was nothing to point in that direction.

The amendments introduced in auxiliary request 1 served to emphasize the differences over the disclosure of D3. In particular, they explicitly excluded any reading of the SAX parser in D3, even if it were implemented as an automaton, onto either of the automata defined in claim 1.

The amendments further introduced with auxiliary request 2 served to emphasize that the automata defined in claim 1 did not function as a SAX parser.

The amendments further introduced with auxiliary requests 3 and 4 served to emphasize that the automata only searched for tags and keywords that were derived from the search terms.

Reasons for the Decision

1. *Background*

- 1.1 The invention concerns searching tagged documents. The application does not explain what "tagged" means, but the term is intended to cover documents written in XML, in which the parts of a document are enclosed in hierarchical tags. Examples might be the following fragments:

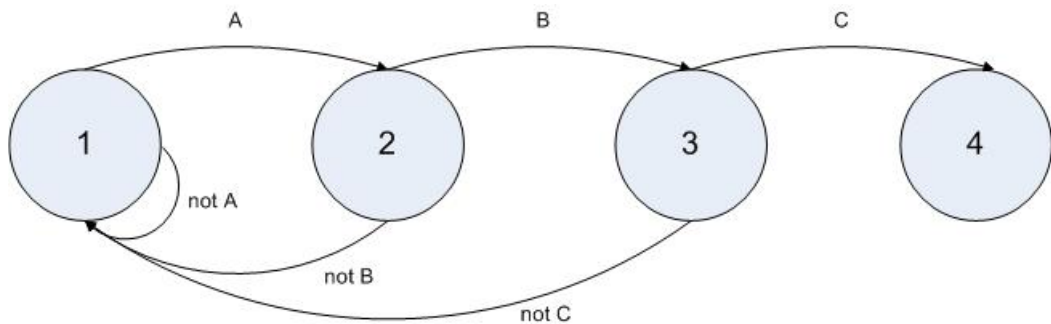
```
<book>
  <author> Ms A. Writer </author>
</book>
```

```
<diary>
  <entry>
    <event> Oral proceedings </event>
    <date> 15 March 2013 </date>
    <author> Ms A. Writer </author>
  </entry>
</diary>
```

- 1.2 The task is to find documents, or parts of documents, that contain a keyword enclosed by tags at the correct place in the hierarchy. Thus, the first example would match a search for `/book/author="Ms A. Writer"`, but the second example would not.
- 1.3 As set out in the introductory part of the description, the prior art methods of finding matching documents involved the analysis and storage of the complete hierarchy of tags for each document. The invention does not take that approach. Instead, it scans the document from the start, looking for tags that match the search query. If it finds a matching tag, it switches to a

different search mode, and looks for a matching keyword.

1.4 According to the invention, the searches for matching tags and (if a matching tag is found) keywords are carried out by automata. That is a term of art, referring to a machine with a finite number of states, in which transitions between states occur depending on some input. An automaton that searches for the string "ABC" might have four states as shown.



State 1 reads its input symbol and does nothing unless it is "A". If the input symbol is "A", then the transition is to state 2. State 2 transitions to state 3, if its input symbol is "B", or else transitions back to state 1. State 3 transitions to state 4 if its input symbol is "C" or else transitions to state 1. As a consequence of these rules, the automaton reaches state 4 if and only if the symbols "A", "B", and "C" occur in that order, consecutively, in the input stream. Arrival in state 4, therefore, indicates that the string "ABC" has been found.

1.5 The present invention uses two automata. The first looks for tags, the second for keywords within a matching tag, if one has been found.

2. *Main request*

2.1 Claim 1 is somewhat unclear, but it is common ground that it seeks to define an apparatus which comprises a "generation device", a "read device", and a "search device". The generation device analyzes a "search condition", creates a first automaton to search for relevant tags and a second automaton to search for relevant keywords, and create information for managing states. The read device reads document data and provides a stream of symbols to the search device. The search device switches between searching for tags and searching for keywords. The switch from tag to keyword search happens when the state management information indicates that the current state is one in which the switch is to occur. The claim does not explicitly state that the searches are actually carried out by the automata, but the Board understands that to be the intention.

2.2 The claim is not limited so searching in XML documents. The document is defined simply as "structured using a tag." The Board sees this as broad enough to cover a book with a contents page, and introduction, some chapters, and an index. Its tags would be inscriptions such as "Introduction", or "Chapter 3." The Board notes, however, that the same book could be written in XML, and the tags would then take the form "<Introduction>" and "<Chapter id=3>," so that the fact that claim 1 is not limited to XML documents makes no difference.

2.3 The Board sees the following method as underlying the apparatus defined in claim 1.

- Scan the document from the beginning until a relevant

tag is found.

- If a relevant tag is found, scan its contents until a relevant keyword is found or the end tag is reached.

- If both a relevant tag and a relevant keyword are found, indicate the fact.

2.4 That underlying method is not technical. It might be no more than a person scanning a book, to see if it has a chapter which talks of constitutional law, for example; or to see whether "constitutional law" occurs in the index. In the Board's view, the method would remain non-technical, even if the book were encoded in XML. That would be a different technical substrate from the paper and ink that underlie a conventional book, but the method operates independently of the technical substrate. It is the contents of the book that are searched, not the physical encoding of those contents.

2.5 The implementation of that underlying method involves, according to claim 1, two automata. One interpretation of the term "automaton" is as a mathematical model of computation, just as Turing machines are. That is not an interpretation that fits an invention in which an automaton is actually used to perform a search. Rather, as the appellant has stated, the correct interpretation is as a computer program. The automata produced by the generation device are programs.

2.6 Programs as such are not regarded as inventions under Article 52(2)(c) EPC. However, if a claim feature relating to a program has some technical effect beyond what is inherent in running a program at all, then there is a technical contribution, and the feature may contribute to inventive step. It is necessary,

therefore, to consider whether the automata can contribute to inventive step in the present case.

2.7 In general, a program may bring about effects outside the computer. That would be the case, for example, if a program controlled the flight of an aeroplane. In the present case, the only external effect (if it is external at all) lies in the information that a document does, or does not, match a search condition. To continue the example above, the output of the search result might be that books 2 and 5 each have a chapter that mentions "constitutional law." The Board cannot see anything technical in that.

2.8 The appellant, however, has argued that there is a technical effect within the computer itself, in terms of the efficient use of computer resources. According to the appellant's arguments, there are two reasons for that. Firstly, many documents can be rejected without ever having to look at keywords, because the tags do not match. Secondly, the second (keyword) automaton is not loaded into working memory when not in use, so that, if matching tags are rare, it need hardly ever be loaded in memory at all. Furthermore, when not in use, the second automaton is removed from memory, so that resources are freed.

2.9 As to the first argument, it is an effect due to the method of searching, not to the method of programming. If a person is looking for books with the keyword "constitutional law" in a chapter titled "Written and unwritten constitutions", then the non-technical method set out above would reject all books that do not have a chapter with that title, without ever looking at the keywords "constitutional law" at all. Therefore, this

is not a technical effect.

2.10 The appellant's second argument relies on a reading of claim 1 such that the second automaton is loaded in memory when it is used, and unloaded when it is not. The appellant's argument is that this reading can be derived from the term "switching" used in the claim, and from the fact that Figure 19 of the application shows memory 1902. In the Board's view, this argument loads the term "switch" with a meaning it cannot bear. Claim 1 does not mention what is and what is not loaded into memory. Nor is there anything in the application as filed that mentions when the second automaton is loaded into or removed from memory. Nor does the Board see any evidence that the efficiency in memory usage is actually obtained. The application does not speak of the problem of memory usage, and the appellant has not provided any evidence such as an analysis or measurement of memory usage with and without the invention. The Board, therefore, rejects the appellant's second argument.

2.11 In conclusion, the Board sees a technical effect neither outside nor inside the computer. As a result, the automata do not contribute to inventive step.

2.12 The underlying method of searching, and the use of automata to perform the search, are, therefore, regarded as non-technical features of claim 1. When assessing inventive step, it is legitimate to ask whether an invention could amount to an automation of an underlying non-technical method. If it does, then the question of inventive step amounts to asking whether, given the method, the particular automation claimed would have been an obvious one.

- 2.13 From that point of view, in order to arrive at the apparatus defined in claim 1, the skilled person would have to provide the "generation device", the "read device" and the "search device" as set out in claim 1.
- 2.14 In the Board's view, the skilled person, faced with the technical problem of automating the non-technical method would have had no choice but to provide some means of analysing the search condition and defining suitable automata. She would, equally, have had no choice about the provision of a device that could read documents. Finally, she would have had no choice but to provide some means of searching for tags, and of switching (if a relevant tag is found) to a search for keywords. That is, the apparatus defined in claim 1 defines exactly those technical features that are necessary for any implementation of the non-technical method.
- 2.15 Such implementations would have been obvious, providing the skilled person knew of suitable means. In the present case, the Board has no doubt that a general-purpose computer could be used.
- 2.16 The Board notes that the appellant's arguments regarding the SAX parser of D3 do not affect this analysis, because it does not rely on the disclosure of D3 or of any other document.
- 2.17 The Board, therefore, cannot allow the main request.
3. *Auxiliary request 1*
- 3.1 Claim 1 according to this request, compared with the main request, comprises a number of added terms.

- 3.2 With regard to the generation device, the amendments spell out what was already inherent in the underlying non-technical method: the automata are defined on the basis of information extracted from the search condition. Similarly, the additional text in the final paragraph serves only to spell out that the states of the automata relate to what is being sought and that the switch from tag search to keyword search happens when it is supposed to.
- 3.3 The same cannot be said of the amendments to the definition of the search device. According to this request, the inputs to the automata are document data read character by character. That is not something inherent in the non-technical method that underlies claim 1 according to the main request.
- 3.4 However, the Board views the reading of a document character by character as non-technical. It is one of the ways that people do read documents (eg one drafted in an unfamiliar language). Thus, this amendment to claim 1 arises as a consequence of an amendment to the underlying non-technical method. Once the non-technical method has been amended, it is inevitable that the search device must take its input character by character, because that is just what the amended non-technical method, which the skilled person is automating, requires.
- 3.5 In conclusion, the amendments are not such as could change the situation with regard to inventive step. Auxiliary request 1 is, therefore, not allowable.

4. *Auxiliary request 2*

4.1 This request adds a substantial amount of text to the end of claim 1. In effect, however, the additional text amounts to a specification that it is the first (tag) automaton that actually searches for matching tags, and the second (keyword) automaton that actually searches for matching keywords.

4.2 As stated in point 2.1, above, the main request does not clearly specify that it is the automata that perform the search, but the Board understands that they do. The assessment of the main request, therefore, already takes account of the amendment, and the result, as far as inventive step is concerned, is the same as for auxiliary request 1.

4.3 Therefore, auxiliary request 2 cannot be allowed.

5. *Auxiliary request 3*

5.1 The additional text in claim 1 is to the effect that the character strings for which the automata search are derived from the search condition. That is, the automata do not search for something that is not part of the search condition.

5.2 This too is (at most) an amendment to the underlying non-technical method of searching. As such, it does not affect the outcome of the assessment of inventive step.

5.3 Auxiliary request 3 cannot be allowed.

6. *Auxiliary request 4*

6.1 This request was filed during oral proceedings before the Board. The amended text is only concerned with the underlying non-technical method of searching, so that the request is not clearly allowable.

6.2 The Board, therefore, does not admit the request (Article 13 RPBA).

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chairman:



T. Buschek

S. Wibergh

Decision electronically authenticated