

**Internal distribution code:**

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

**Datasheet for the decision  
of 11 November 2016**

**Case Number:** T 2439/11 - 3.5.07

**Application Number:** 07856906.8

**Publication Number:** 2126750

**IPC:** G06F17/30

**Language of the proceedings:** EN

**Title of invention:**

Method for classifying web pages and organising corresponding contents

**Applicant:**

FacilityLive Opco S.r.l.

**Headword:**

Web-page classification/FACILITYLIVE OPCO

**Relevant legal provisions:**

EPC Art. 56, 123(2)

**Keyword:**

Amendments - added subject-matter (yes)  
Inventive step - (no)

**Decisions cited:**

T 1358/09

**Catchword:**



**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 2439/11 - 3.5.07

**D E C I S I O N**  
**of Technical Board of Appeal 3.5.07**  
**of 11 November 2016**

**Appellant:** FacilityLive Opco S.r.l.  
(Applicant) Via Strada Nuova 86  
27100 Pavia (IT)

**Representative:** Botti, Mario  
Botti & Ferrari S.r.l.  
Via Cappellini, 11  
20124 Milano (IT)

**Decision under appeal:** Decision of the Examining Division of the  
European Patent Office posted on 24 May 2011  
refusing European patent application No.  
07856906.8 pursuant to Article 97(2) EPC.

**Composition of the Board:**

**Chairman** R. Moufang  
**Members:** R. de Man  
M. Jaedicke

## Summary of Facts and Submissions

I. The applicant, which at the time was Mouldtec Ontwerpen B.V., appealed against the decision of the Examining Division refusing European patent application No. 07856906.8, filed as international application PCT/EP2007/011183 and published as WO 2008/074486.

II. Subsequently, the application was transferred twice, first to FacilityLive S.r.l. registered with effect from 31 July 2012 and then to FacilityLive Opco S.r.l. (appellant) registered with effect from 11 April 2014.

III. The Examining Division refused the application for lack of inventive step in the subject-matter of claim 1 of a sole substantive request in view of the following document:

D2: US 2004/186828 A1, 23 September 2004.

IV. With the statement of grounds of appeal, the then appellant maintained the claims considered in the decision under appeal as its sole substantive request.

V. In a communication under Article 15(1) RPBA following a summons to oral proceedings, the Board introduced the following document, corresponding to one of the documents cited in the international search report:

D3: Attardi G. et al.: "Automatic Web Page Categorization by Link and Context Analysis", Pre-proceedings of the THAI-ETIS European symposium, 21-22 June 1999, pages 1-15, retrieved from <URL:<http://puma.isti.cnr.it/linkdoc.php?idauth=21&idcol=1&icode=1999->

A2-003&authority=cnr.iei&collection=cnr.isti&langver=en>.

The Board raised an objection under Article 123(2) EPC against claim 1, expressed the preliminary view that the Examining Division's reasoning starting from document D2 was not convincing but that the subject-matter of claim 1 nevertheless lacked inventive step over document D3 and drew attention to a number of possible clarity problems in claim 1.

- VI. By letter of 28 October 2016, the appellant withdrew its request for oral proceedings. It made no substantive comments on the Board's communication.
- VII. Oral proceedings were held on 11 November 2016 in the appellant's absence. At the end of the oral proceedings, the chairman pronounced the Board's decision.
- VIII. The appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of the claims of the sole substantive request.
- IX. Claim 1 of the sole substantive request reads as follows:

"Method for classifying Web pages and organizing the corresponding contents of the type comprising

- a step o [sic] defining lemmas,
- a step of recording a predetermined number of Internet addresses of Web pages associated to said lemmas, comprising executions of a plurality of

automatic recording processes of said plurality of Internet addresses,

- a step of selecting and setting a corresponding pertinence value to said plurality of Internet addresses, characterized in that

said selection step sets said corresponding pertinence value to a pertinence value proportional to the recordings of said Internet address in said recording step and selects the Internet addresses with pertinence value greater than a predetermined threshold value,

the method furthermore comprising

- a reduction step executed by an identification script which, from among said Internet addresses with pertinence value greater than said threshold value, reduces said Internet addresses to the Internet addresses meeting one or more essentiality criteria, said essentiality criteria including eliminating Web pages coming from an identical domain from the same lemma, and
- a validation step for validating a subset of said Internet addresses meeting the essentiality criteria, said validation step comprising a human action."

X. The appellant's arguments as relevant to this decision are discussed in detail below.

## Reasons for the Decision

1. The appeal complies with the provisions referred to in Rule 101 EPC and is therefore admissible.
2. *The invention*
  - 2.1 The background section of the application refers to known web-based search engines (referred to in the application as "identification scripts or programs") such as the Google search engine shown in Figures 1, 1a and 1b. Such a search engine automatically selects and orders search results on the basis of the results of a classification method. One such classification method is Google's PageRank algorithm, which assigns an "importance value" to each web page.
  - 2.2 The application identifies the following limitations of known classification methods, illustrated by the search results shown in Figures 1, 1a and 1b:
    - some of the top-ranked results are redundant in that they refer to web pages on the same web site (page 4, line 30, to page 5, line 7);
    - some high-ranked web pages are not pertinent to the information searched by the user (page 5, lines 8 to 22);
    - a number of search hits is reported that does not correspond to the number of pages that can actually be retrieved by the user (page 5, line 27, to page 6, line 12); and
    - search results may include web pages with unreliable information (page 6, lines 13 to 25).
  - 2.3 The application therefore sets out to provide a web page classification method that allows a search engine

("identification script or program") to provide "results which are free of information noise, pertinent to the search criterion set by the user, not redundant for repetition of Web pages, and reliable for their content" (page 7, lines 9 to 16).

2.4 According to the invention as defined by claim 1, first a number of "lemmas" are defined, presumably by a human. As explained on page 10, lines 5 to 9, the term lemma is used to indicate headwords and other forms of abstract units of morphological analysis in linguistics, each abstract unit roughly corresponding to a set of words that are different forms of the same word.

Then a number of "automatic recording processes" are executed to record a "predetermined" number of Internet addresses associated with those lemmas.

A pertinence value is assigned "to said plurality of Internet addresses". This value is chosen to be "proportional to the recordings of said Internet address". The Board understands this as meaning that a pertinence value is assigned to each internet address and that this value proportionally reflects the number of times the address was recorded in connection with one of the lemmas by the plurality of automatic recording processes.

All internet addresses with a pertinence value greater than a predetermined threshold value are "selected".

In a reduction step, the selected internet addresses are "reduced" to those meeting one or more essentiality criteria "including eliminating Web pages coming from an identical domain for the same lemma".

In a validation step, a human then "validates" the Internet addresses remaining after the reduction step.

3. *Added subject-matter*

- 3.1 The method of claim 1 includes a "reduction step executed by an identification script". Although in the claim's context the term "identification script" could be understood simply as referring to a script that performs the functionality specified by the "reduction step" feature, the description of the application uses the term "identification script" in a specific sense.

According to the description on page 1, line 26, to page 2, line 13, identification scripts are scripts for identifying web pages on the basis of a search criterion input by a user. Such scripts are generally accessible through a graphical user interface, for example a browser, which comprises an "insertion portion" for entering the search criterion and a button for activating the script. The description on page 2, lines 8 to 13, further explains that an identification script may be supported by a classification method that assigns an "importance" to each web page. Execution of an identification script provides the user with a list of web pages ordered by importance value as established by the classification method.

Thus, an "identification script" is the component of an internet search engine that takes the search criteria entered by a user as input and searches a collection of (classified) web pages for pages matching the search criteria.

This understanding of the term "identification script" is confirmed by other passages of the description of the application (see in particular page 2, line 30, to page 4, line 7; page 4, line 30, to page 5, line 22; page 5, line 27, to page 6, line 12; and page 6, line 26, to page 7, line 16).

- 3.2 The feature specifying that the reduction step is executed by an identification script was not present in the originally filed claims, and the Board is unable to find support for it in the originally filed description. According to the appellant's letter of 11 April 2011, it is based on the description on page 8, lines 22 to 24, page 15, lines 25 to 28, and page 15, first and second paragraphs.
- 3.3 The Board, however, understands the passage on page 8, lines 22 to 27, as stating merely that the classification method of the application enables a search engine ("identification script or program") to provide "a result pertaining to ...". This passage does not disclose that the classification method itself, or certain steps of it, are carried out by an internet search engine.
- 3.4 Similarly, the description on page 15 discloses a "reduction step" as part of the classification method. Because of this reduction step, which results "in a third, further refined information layer or repository" (page 15, lines 22 to 24), "an identification script or program executed on the third information layer or repository is capable of identifying internet addresses with high pertinence probability" (page 15, lines 25 to 28; underlining added by the Board). Thus, here too the term "identification script" refers to (the component of) an

internet search engine that does not carry out steps of the classification method but merely uses the results of the classification when performing searches. This is further confirmed by the passage on page 15, line 29, to page 16, line 6.

3.5 Hence, the subject-matter of claim 1 extends beyond the content of the application as filed, contrary to Article 123(2) EPC.

4. *Inventive step*

4.1 Since the objection of added subject-matter could have been overcome in a straightforward manner, namely by deleting the words "executed by an identification script", the Board considers it appropriate in the present case to base its decision on lack of inventive step as well.

4.2 Although the offending feature has no basis in the application as filed, when assessing inventive step the Board cannot ignore it, because it is bound by the wording of the claims as submitted by the appellant (Article 113(2) EPC).

Since it is apparent from the application as a whole that the claimed reduction step is part of a classification method that is performed before a search criterion is entered by a user and a corresponding search is carried out, the Board considers that "a reduction step executed by an identification script" is to be understood as meaning that the identification script of the method of claim 1 includes not only instructions implementing the actual search but also a (separate) set of instructions for carrying out the reduction step.

4.3 Document D3 relates to automatic web page classification (see abstract). It describes on page 3, lines 22 to 42, a known web search engine for searching a collection of documents that have been automatically pre-classified according to subject on the basis of a subject hierarchy of 20 000 handcrafted terms. On page 3, line 43, to page 4, line 8, the document discusses further examples of search engines that use automatic document classification techniques for organising documents into categories.

Hence, document D3 discloses a method for classifying a collection of web pages and organising the corresponding contents. This method comprises a step of defining (handcrafted) terms and a step of recording internet addresses of web pages associated with the terms in the sense of assigning web pages to terms. The subject-matter of claim 1 differs from this prior-art method in the following additional features:

- (a) documents are classified on the basis of "lemmas" rather than terms (cf. point 2.4 above);
- (b) a "predetermined" number of internet addresses is recorded;
- (c) recording comprises executing a plurality of automatic recording processes;
- (d) pertinence values are assigned "proportional to the recordings" of the internet addresses;
- (e) (only) the internet addresses with pertinence value greater than a predetermined threshold value are determined;
- (f) in a "reduction step", web pages coming from an identical domain for the same lemma (as other web pages in the selection) are eliminated;

(g) the remaining web pages/internet addresses are manually "validated"; and

(h) the reduction step (f) is executed by an identification script.

These features largely correspond to the distinguishing features which are identified in the decision under appeal with respect to document D2 and on which the appellant based its arguments in support of inventive step in the statement of grounds of appeal.

4.4 According to the appellant, the distinguishing features identified in the decision had a synergistic effect consisting in classifying and organising web pages more quickly compared to prior-art methods.

Given that the claimed method in any case returns different results than the method of document D3 or any other prior-art method, the fact that the claimed method runs faster than known methods - if that is indeed the case - is in itself not very meaningful. In its communication the Board therefore suggested that the appellant's argument was to be understood as saying that the distinguishing features interact to efficiently classify web pages with good results, i.e. that it achieved a good compromise between speed of execution and quality of search results. Indeed, the description presents the problem to be solved as that of obtaining pertinent and reliable results which are free of information noise (cf. point 2.3 above).

4.5 In the Board's view, however, such an effect does not qualify as a technical effect. In the context of the present application, the "quality" of a particular classification of web pages is a cognitive matter and, therefore, not a technical issue. The question whether

or not a web page about the Leonardo da Vinci airport is to be classified under the lemma "Leonardo da Vinci", to take an example from the description of the application, is answered on the basis of cognitive considerations. See in this respect also decision T 1358/09 of 21 November 2014, reasons 5.2.

4.6 Nevertheless, it still remains to be determined whether an inventive step is present in technical considerations underlying the individual steps or combinations of steps or in their technical implementation.

4.7 In the present case, feature (c), interpreted in the light of the description on page 11, line 19, to page 12, line 4, and on page 13, lines 16 to 28, refers to the use of technical means in the form of spidering processes and "meta-search engine functions" (i.e. queries of other search engines) for the purpose of collecting internet addresses.

However, both the use of spidering processes and the use of meta-search engine functions for collecting internet addresses were known in the art at the priority date of the application (see, for spidering, document D2, paragraphs [0053] to [0057], and document D3, page 5, lines 1 to 11, and page 9, lines 8 to 11, and, for meta-search engine functions, document D2, paragraphs [0044] to [0048], and the description of the application on page 13, lines 16 to 21 ("as known")). Feature (c) is therefore obvious.

4.8 Feature (b) arguably reflects the technical consideration that limiting the number of internet addresses to be recorded limits resource usage, but that consideration is obvious and trivial to implement.

4.9 Features (a), (d), (e), (f) and (g) reflect only non-technical considerations on how to improve the quality of the resulting classification. The technical implementation of features (a), (d), (e) and (f) on a computer is straightforward and not further described in the application. Implementing feature (f) in accordance with feature (h) as part of an "identification script" reflects an obvious implementation choice not leading to a surprising technical effect.

The manual validation step of feature (g) is presumably to be facilitated by a user interface, but the implementation of such an interface must likewise be considered to fall within the abilities of the skilled person. The description of the present application again gives no implementation details.

4.10 Thus, although the particular claimed combination of features is novel, since no inventive step can be seen in technical considerations underlying the individual steps or combinations of steps or in their technical implementation, the subject-matter of claim 1 lacks inventive step (Articles 52(1) and 56 EPC).

5. Since the sole substantive request cannot be allowed, 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