BESCHWERDEKAMMERN DES EUROPÄISCHEN PATENTAMTS	BOARDS OF APPEAL OF THE EUROPEAN PATENT OFFICE	CHAMBRES DE RECOURS DE L'OFFICE EUROPÉEN DES BREVETS

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 March 2021

Case Number: T 0189/19 - 3.5.07

Application Number: 12817909.0

Publication Number: 2737420

IPC: G06F17/30

Language of the proceedings: EN

Title of invention:
Utilization of features extracted from structured documents to improve search relevance

Applicant:
Microsoft Technology Licensing, LLC

Headword:
Ranking semi-structured documents/MICROSOFT TECHNOLOGY LICENSING

Relevant legal provisions:
EPC Art. 56, 84, 123(2)
RPBA Art. 12(4)
RPBA 2020 Art. 13(2)

Keyword:
Inventive step - main request (no)
Amendments - discretion to admit requests not admitted by the
examining division - first auxiliary request (yes)
Added subject-matter - first auxiliary request (yes)
Amendment after summons - exceptional circumstances - second
auxiliary request (no)
Claims - clarity - third auxiliary request (no)

Decisions cited:
T 1816/11, T 1159/13, T 2343/13, T 0336/14

Case Number: **T 0189/19 - 3.5.07**

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

Appellant: (Applicant)	Microsoft Technology Licensing, LLC One Microsoft Way Redmond, WA 98052 (US)
Representative:	Goddar, Heinz J. Boehmert & Boehmert Anwaltspartnerschaft mbB Pettenkoferstrasse 22 80336 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 14 August 2018 refusing European patent application No. 12817909.0 pursuant to Article 97(2) EPC**

Composition of the Board:

Chair	M. Jaedicke
Members:	R. de Man
	E. Mille

Summary of Facts and Submissions

I.   The applicant (appellant) appealed against the decision
         of the examining division refusing European patent
         application No. 12817909.0, published as international
         application WO 2013/016288.

II.   The contested decision cited, *inter alia,* the following
         documents:

         D1: US 2009/125529 A1, 14 May 2009;
         D3: US 2009/234848 A1, 17 September 2009.

         The examining division decided that the subject-matter
         of claim 1 of the main request infringed Article 123(2)
         EPC. Claim 1 of the first auxiliary request was not
         clear and its subject-matter did not involve an
         inventive step over document D1 in combination with
         document D3. Claim 1 of the second auxiliary request
         was not clear and its subject-matter did not involve an
         inventive step over document D1.

         In a section of the decision titled "Further remarks",
         the examining division explained that auxiliary
         requests 1 and 2 submitted by the applicant on
         5 June 2018 and "the auxiliary request filed during the
         oral proceedings at 14:35" had not been admitted into
         the proceedings under Rule 137(3) EPC.

III.   With its statement of grounds of appeal, the appellant
         filed a main request and first and second auxiliary
         requests. The main request corresponded to the first
         auxiliary request considered in the decision under
         appeal. The first auxiliary request corresponded to the

"second auxiliary request" filed at 14.35h during the oral proceedings before the examining division. The second auxiliary request corresponded to the "revised second auxiliary request" filed at 16.00h in the oral proceedings before the examining division and considered in the decision under appeal as second auxiliary request.

IV. In a communication accompanying the summons to oral proceedings, the board expressed the preliminary view that claim 1 of the main request was not clear and that its subject-matter lacked inventive step when starting from document D1 as closest prior art. With respect to the first auxiliary request, it expressed doubts as to its admittance into the appeal proceedings and raised a number of clarity and added-matter objections. With respect to the second auxiliary request, it raised a number of clarity objections and expressed its agreement with the inventive-step analysis contained in the contested decision.

V. With a letter dated 10 February 2021, the appellant replaced its main request and second auxiliary request with a revised main request and a revised second auxiliary request.

VI. During the oral proceedings, which took place on 10 March 2021, the appellant filed a new second auxiliary request and maintained its previous second auxiliary request as a third auxiliary request. At the end of the oral proceedings, the Chair pronounced the board's decision.

VII. 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 main request filed with the letter

of 10 February 2021 or, in the alternative, of one of
the first auxiliary request filed with the statement of
grounds of appeal, the second auxiliary request filed
in the oral proceedings before the board, and the third
auxiliary request filed (as "revised second auxiliary
request") with the letter of 10 February 2021.

VIII.   Claim 1 of the main request reads as follows:

"A method (1100) of outputting a ranked list of
documents responsive to receipt of a query, the method
comprising:
at a general purpose search engine, receiving (1104)
the query from a user to retrieve at least one document
from a plurality of documents in a data store, wherein
the plurality of documents comprises a plurality of
semi-structured web pages and the at least one document
is indexed by the general purpose search engine;
generating a ranked list of documents based at least in
part upon the query, wherein the generating comprises
selectively positioning at least one semi-structured
web page of a plurality of semi-structured web pages at
a particular position in the ranked list of documents
based at least in part upon a value of a feature that
is extracted at a predetermined location in the at
least one semi-structured web page, wherein the
position of the at least one semi-structured web page
in the ranked list of documents is independent of any
correlation between content of the query and the value
of the feature; and
causing a processor to output (1106) the ranked list of
documents comprising the at least one semi-structured
web page to the user,
characterized in that the method further comprises
automatically extracting features from the plurality of
semi-structured web pages, wherein the feature is

automatically extracted from the at least one semi-structured web page at the predetermined location."

IX.   Claim 1 of the first auxiliary request reads as follows:

"A method (1100) of preprocessing and outputting a ranked list of documents responsive to receipt of a query, the method comprising:
at the general purpose search engine, receiving (1104) the query from a user to retrieve at least one semi-structured webpage of a plurality of semi-structured web pages that is indexed by the general purpose search engine;
generating a ranked list of documents based at least in part upon the query;
selectively positioning the at least one semi-structured web page of the plurality of semi-structured web pages at a particular position in the ranked list of documents based at least in part upon a value of a feature that is extracted at a learned location in the at least one semi-structured web page, wherein the position of the at least one semi-structured web page in the ranked list of documents is independent of any correlation between content of the query and the value of the feature; and
causing a processor to output (1106) the ranked list of documents comprising the at least one semi-structured web page to the user,
characterized in that the method further comprises a preprocessing step of automatically extracting features that are consistent across semi-structured web pages, wherein identities of the features are learned through analysis of query click logs of the general purpose search engine, from the plurality of semi-structured web pages, wherein the feature is automatically

extracted from the at least one semi-structured web page at the learned location."

X.   Claim 1 of the second auxiliary request differs from claim 1 of the first auxiliary request in that the text "and that have some bearing on the relevance of the semi-structured web pages to the query" has been inserted after "a preprocessing step of automatically extracting features that are consistent across semi-structured web pages".

XI.   Claim 1 of the third auxiliary request reads as follows:

"A method (1100) of preprocessing and outputting a ranked list of documents responsive to receipt of a query, the method comprising:
learning a wrapper that is used to automatically identify features and extract feature values from a plurality of semi-structured web pages;
learning a scoring function that assigns scores to the features based at least in part upon values of the features and query/click logs of a general purpose search engine;
assigning scores to the features extracted from the plurality of semi-structured web pages based on the learned scoring function;
at the general purpose search engine, receiving (1104) the query from a user to retrieve at least one document that is indexed by the general purpose search engine, the at least one document comprising at least one semi-structured webpage of the plurality of semi-structured web pages;
generating a ranked list of documents based at least in part upon the query;

selectively positioning the at least one semi-structured web page of the plurality of semi-structured web pages at a particular position in the ranked list of documents based at least in part upon a value of a feature that is automatically extracted at a predetermined location in the at least one semi-structured web page, wherein the position of the at least one semi-structured web page in the ranked list of documents is independent of any correlation between content of the query and the value of the feature; and causing a processor to output (1106) the ranked list of documents comprising the at least one semi-structured web page to the user."

XII.   The appellant's arguments, where relevant to the decision, are discussed in detail below.

Reasons for the Decision

1.   *The application*

1.1   The application relates to ranking "semi-structured web pages" retrieved in response to a search query.

1.2   According to paragraph [0027] of the published application, "a semi-structured web page is a web page with content that fails to conform to the structure of a relational database, but nevertheless comprises a pattern that is consistent across a plurality of other semi-structured web pages". Web pages conforming to the same "pattern" include common features in corresponding positions such as an image, a title, reviews, a number of views, a number of comments (paragraphs [0027], [0029], [0030]; Figures 2 and 3).

1.3 The application essentially proposes ranking semi-structured web pages on the basis of the content ("value") of such a feature (paragraph [0033]). For example, a higher number of positive reviews may cause a web page to be ranked higher (paragraph [0034]).

Main request

2. *Admission in the appeal proceedings*

The current main request is based on the previous main request filed with the statement of grounds of appeal and includes amendments addressing a number of clarity objections raised for the first time in the board's communication. Since the amendments were filed with the letter of 10 February 2021 and thus at the earliest opportunity, the admission of the main request into the appeal proceedings is justified by an exceptional circumstance as required by Article 13(2) RPBA 2020.

3. *The board's interpretation of claim 1*

3.1 Claim 1 of the main request is directed to a method of outputting a ranked list of documents in response to a search query.

3.2 First, a general-purpose search engine receives the query from a user and generates a ranked list of search results retrieved from a plurality of documents indexed by the search engine. The plurality of documents comprises a plurality of semi-structured web pages.

3.3 Generating the ranked list of search results involves positioning at least one retrieved semi-structured web page at a particular position in the ranked list. This position is determined, at least in part, on the basis

of the value of a "feature" at a predetermined location
in the web page, whereby no account is being taken of
any correlation between the feature's value and the
query's content.

3.4   The ranked list is output by a processor to the user.

3.5   The claim further includes the feature "automatically
        extracting features from the plurality of semi-
        structured web pages" and specifies that "the feature
        is automatically extracted from the at least semi-
        structured web page at the predetermined location".
        Since the claim also refers to "a value of a feature
        that is extracted at a predetermined location", the
        skilled person reading the claim understands that what
        is (automatically) extracted are the values of
        features.

4.   *Inventive step*

4.1   Document D1 relates to extracting attributes from web
        pages (see abstract). It describes, in paragraphs
        [0079] to [0082], an internet search engine that crawls
        the World Wide Web to index web pages. For pages that
        include job descriptions, information such as job
        title, job location and required experience is
        extracted from the page and used to index the page in
        the search index. This information is extracted with
        the help of extraction templates.

4.2   Extraction templates are automatically created from
        training documents by a process described in paragraphs
        [0083] to [0159] (see in particular paragraphs [0083]
        to [0086]). The extraction templates include
        information on the location of features ("attributes")
        in a document's DOM tree, which is used to

automatically extract the feature's value (paragraphs
[0160] to [0163]).

4.3   Web pages matching a particular extraction template are
         "semi-structured" web pages within the meaning of the
         present application (see point 1.2 above).

4.4   Document D1 further discloses, in paragraph [0011],
         that the search engine interface of a search engine
         allows users to specify a search query by means of
         keywords and, in response to the query, displays the
         search results to the user, typically as a ranked list.

4.5   Hence, the subject-matter of claim 1 differs from the
         disclosure of document D1 in that the position of a
         semi-structured web page in the ranked list of
         documents is based, at least in part, on the value of a
         feature of the document, whereby no account is being
         taken of any correlation between the feature's value
         and the query's content.

4.6   The appellant argued that this distinguishing feature
         led to more accurate search results when a database of
         documents was searched. Since accessing and searching
         databases was commonly considered to be a technical
         problem, more accurate search results represented a
         technical effect.

4.7   The board will leave aside the question whether the
         distinguishing feature, which specifies neither what
         kind of feature value is being taken into account nor
         how the value affects the ranking of a web page,
         plausibly improves the accuracy of search results over
         the whole scope of the claim and will focus instead on
         the specific example described in paragraph [0007] of
         the application, which suggests that a web page that

contains a greater number of positive reviews may be positioned in the search results above a web page that contains fewer positive reviews or more negative reviews. Although the board is willing to accept that this leads to search results which are more relevant to the typical user, this relates to the subjective appreciation of the cognitive content of the search results and is not a technical improvement. Indeed, the insight that a greater number of positive reviews indicates a greater relevance to the user is not one that belongs to a technical field.

4.8   The appellant also argued that the invention involved a continued and guided human-machine interaction process, which was technical according to decision T 336/14.

        However, the distinguishing feature does not relate to human-machine interaction. Any human-machine interaction specified in the claim is already present in document D1.

4.9   Hence, the board does not agree with the appellant that the distinguishing feature achieves a technical effect. The problem to be solved may therefore be formulated as how to modify the disclosure of document D1 so as to base the position of a semi-structured web page in the ranked list of documents, at least in part, on the value of a document feature which has no correlation with the query's content. Since document D1 already discloses extracting values of document features, this problem amounts to a straightforward and thus obvious programming exercise for the skilled person.

4.10   Hence, the subject-matter of claim 1 lacks inventive step (Article 56 EPC).

First auxiliary request

5.   *Admission into the proceedings – Article 12(4)*
        RPBA 2007

5.1   The first auxiliary request corresponds to the "second
        auxiliary request" filed at 14.35h during the oral
        proceedings before the examining division.

5.2   According to the (corrected) minutes of those oral
        proceedings, auxiliary requests 1 and 2 as filed with
        the letter of 5 June 2018 were first discussed in the
        oral proceedings and were then withdrawn when the
        appellant, at 12.15h, filed a new first auxiliary
        request.

        Likewise, it appears that the second auxiliary request
        filed at 14.35h was withdrawn when the appellant, at
        16.00h, filed a "revised second auxiliary request".
        Although point 53 of the minutes states that the
        examining division had "decided" not to admit the
        request, this appears to be merely an unfortunate
        choice of wording. Indeed, point 49 also uses the word
        "decided" in relation to the same request, even though
        the discussion on the request then continued.

5.3   Although the decision under appeal does mention that
        the "auxiliary request filed during the oral
        proceedings at 14:35" had not been admitted into the
        proceedings under Rule 137(3) EPC, it does so in a
        section titled "Further remarks", which also mentions
        that auxiliary requests 1 and 2 submitted with the
        letter of 5 June 2018 had not been admitted under
        Rule 137(3) EPC. In the final "Decision" section, no
        reference is made to any of these requests. Nor are

these requests included in the list of requests on which the decision is based.

Hence, the decision under appeal is not inconsistent with the view that the second auxiliary request filed at 14.35h during the oral proceedings before the examining division, as well as the auxiliary requests 1 and 2 filed with the letter of 5 June 2018, was withdrawn. The withdrawal of a request during the first-instance proceedings has to weigh heavily against the admission of the identical request into the appeal proceedings.

5.4   However, the documents on file also allow for a different interpretation, namely that the examining division, in accordance with the wording of point 53 of the minutes, had indeed decided not to admit the second auxiliary request filed at 14.35h. This would mean that the request could no longer be withdrawn (whether or not that was the intention of the "revised second auxiliary request" filed at 16.00h) and that the examining division incorrectly did not include the request in the list of requests on which the decision was based and, again incorrectly, justified the request's non-admission only under the heading "Further remarks".

At the oral proceedings before the board, the appellant indicated that it had indeed assumed that the examining division had decided not to admit the second auxiliary request filed at 14.35h into the proceedings and that it had not intended to withdraw the request.

5.5   An appellant is responsible for checking the minutes of oral proceedings without delay, in particular to make sure that they correctly reflect what it considers to

be its final requests. In the present case, the
appellant did request a correction of the minutes, but
only to include the second auxiliary request filed at
14.35h as one of the annexes to the minutes.

5.6   Nevertheless, the minutes as they stand do leave doubt
        as to whether the second auxiliary request filed at
        14.35h was withdrawn. And the somewhat curious
        treatment of the request under "Further remarks" in the
        written decision suggests that the examining division
        itself may have been unsure about the request's legal
        status.

        In view of this situation, the board will assume in the
        appellant's favour that the second auxiliary request
        filed at 14.35 was not withdrawn but instead was not
        admitted into the proceedings under Rule 137(3) EPC.

5.7   According to the "Further remarks" section of the
        contested decision, the request was not admitted
        because it was deemed to be *prima facie* non-compliant
        with Articles 123(2) and 84 EPC, i.e. for non-
        compliance with provisions of substantive patent law.
        This means that the board is in no way bound by the
        request's non-admission by the examining division,
        since otherwise the appellant would be deprived of a
        full judicial review of the examining division's
        assessment of substantive aspects of the case (see
        decisions T 1816/11, Reasons 2.6; T 1159/13,
        Reasons 5.4; and T 2343/13, Reasons 7.7).

5.8   Since, moreover, the board can decide on the
        allowability of the first auxiliary request without
        difficulty, it admits the request into the appeal
        proceedings (Article 12(4) RPBA 2007).

6.   *Added subject-matter - Article 123(2) EPC*

6.1   Claim 1 includes the features "a preprocessing step of
          automatically extracting features that are consistent
          across semi-structured web pages, wherein identities of
          the features are learned through analysis of query
          click logs of the general purpose search engine, from
          the plurality of semi-structured web pages".

6.2   According to the appellant, these features are based on
          paragraph [0028] of the published application. The
          first two sentences of this paragraph read as follows:

          "Features that are consistent across semi-structured
          web pages may have some bearing on the relevance of the
          semi-structured web pages to a query set forth by a
          user of a search engine. Identities of such features
          may be learned, for example, through analysis of query
          click logs of a search engine."

6.3   The skilled reader of paragraph [0028] would understand
          that query click logs give information about the
          relevance to the user of search results, and this is
          confirmed by the remainder of the paragraph. Hence, the
          term "such features" in the second sentence refers not
          to all the "[f]eatures that are consistent across semi-
          structured web pages" in the first sentence but only to
          those features which "may have some bearing on the
          relevance of the semi-structured web pages to a query
          set forth by a user of a search engine". At the oral
          proceedings, the appellant agreed with this
          interpretation of paragraph [0028].

          However, claim 1 specifies that "identities of the
          features are learned through analysis of query click

logs", where "the features" are "features that are
consistent across semi-structured web pages".

Paragraph [0028] therefore does not provide a basis for
these features of claim 1.

6.4   The appellant did not contest that none of the other
       passages of the application which mention "query click
       logs" provide a basis for these feature.

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

Second auxiliary request

7.   *Admission into the proceedings - Article 13(2)*
       RPBA 2020

7.1   The second auxiliary request was filed during the oral
       proceedings before the board. Its admission into the
       appeal proceedings is therefore to be assessed under
       Article 13(2) RPBA 2020. This provision stipulates that
       any amendment to the appellant's appeal case made after
       the notification of the summons to oral proceedings is,
       in principle, not to be taken into account unless there
       are exceptional circumstances which have been justified
       with cogent reasons by the appellant.

7.2   The second auxiliary request is intended to overcome
       the objection of added subject-matter raised in respect
       of the first auxiliary request. This objection was
       included in the board's communication, which means that
       the appellant could have filed the second auxiliary
       request already in response to that communication.
       Instead, in its letter of 10 February 2021, it did not

comment on the objection at all but merely stated that it would "present arguments regarding the first auxiliary request during the Oral Proceedings".

7.3   At the oral proceedings, the appellant explained that it had not filed the request earlier because the objection and the manner in which it had understood the objection and how it could be overcome only during the oral proceedings. In view of this exceptional circumstance, the request had to be admitted.

        However, the board's communication contained detailed reasons for the objection which, in the board's judgment, should have allowed the appellant to understand the objection and to consider whether it was necessary to file an amendment.

7.4   Since the board does not see any exceptional circumstances justifying the filing of the second auxiliary request only during the oral proceedings rather than in advance of them, it does not admit the second auxiliary request into the appeal proceedings (Article 13(2) RPBA 2020).

Third auxiliary request

8.   The third auxiliary request is based on the previous second auxiliary request filed with the statement of grounds of appeal and includes amendments addressing a clarity objection raised for the first time in the board's communication. Since the amendments were filed with the letter of 10 February 2021 and thus at the earliest opportunity, the admission of the third auxiliary request into the appeal proceedings is justified by an exceptional circumstance as required by Article 13(2) RPBA 2020.

9.   *Clarity*

9.1   Claim 1 of the third auxiliary request includes a
        number of features specifying that a "scoring function"
        is learned which assigns scores to features.

        However, claim 1 does not define any use of the
        assigned scores. In particular, it does not express
        that the scores are somehow used in the process of
        ranking the documents retrieved by the general-purpose
        search engine. The claim merely states that the
        position of a semi-structured web page in the ranked
        list of documents is based "at least in part upon a
        value of a feature that is automatically extracted at a
        predetermined location in the at least one semi-
        structured web page".

9.2   At the oral proceedings, the appellant argued that,
        since the claim specified that scores were assigned to
        features, the wording "based ... upon a value of a
        feature" in the context of claim 1 meant that the value
        was a value weighted by the score assigned to the
        feature.

        The board cannot agree that this interpretation is
        implied by the wording of the claim, since claim 1 does
        not make this connection between "based ... upon a
        value of a feature" and the assigned scores. Article 84
        EPC requires the claims to be clear from their wording
        alone, so if the position of a web page was to be based
        on a feature value weighted by the feature's score,
        this should have been expressed in the claim.

        Moreover, according to paragraph [0054] of the
        published application, the assigned scores are used to

train a "ranker component 118 such that the ranker
component 118 takes into consideration values of the
features that have been described above when ranking
search results that are output responsive to receipt of
a user query". Hence, the ranker component 118, once it
has been trained on the basis of feature scores,
determines a position of a web page on the basis of
feature values, not on the basis of weighted feature
values. The application therefore does not support the
interpretation proposed by the appellant.

9.3   As claim 1 does not specify how the learned and
        assigned scores are linked to the other features of the
        claim, claim 1 is not clear (Article 84 EPC).

Conclusion

10.   Since none of the requests admitted into the appeal
        proceedings is allowable, 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