

**Internal distribution code:**

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

**Datasheet for the decision  
of 5 October 2016**

**Case Number:** T 1129/12 - 3.5.07

**Application Number:** 06100200.2

**Publication Number:** 1679613

**IPC:** G06F17/21

**Language of the proceedings:** EN

**Title of invention:**

Method and apparatus for detecting pagination constructs including a header and a footer in legacy documents

**Applicant:**

Xerox Corporation

**Headword:**

Header and footer detection/XEROX

**Relevant legal provisions:**

EPC Art. 56, 84

**Keyword:**

Inventive step - (no) - main request and auxiliary requests I-III  
Lack of clarity and support in the description - auxiliary request IV

**Decisions cited:**

**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 1129/12 - 3.5.07

**D E C I S I O N**  
**of Technical Board of Appeal 3.5.07**  
**of 5 October 2016**

**Appellant:** Xerox Corporation  
(Applicant) Xerox Square - 20A,  
100 Clinton Avenue South  
Rochester,  
New York 14644 (US)

**Representative:** Grünecker Patent- und Rechtsanwälte  
PartG mbB  
Leopoldstraße 4  
80802 München (DE)

**Decision under appeal:** **Decision of the Examining Division of the  
European Patent Office posted on 18 November  
2011 refusing European patent application No.  
06100200.2 pursuant to Article 97(2) EPC.**

**Composition of the Board:**

**Chairman** R. Moufang  
**Members:** M. Rognoni  
P. San-Bento Furtado

## Summary of Facts and Submissions

I. The applicant (appellant) appealed against the decision of the Examining Division to refuse the European patent application No. 06100200.2.

II. In the contested decision, the Examining Division held, *inter alia*, that the subject-matter of claim 1 according to the main request and auxiliary request I, both filed with letter dated 29 September 2011, did not involve an inventive step having regard to the following document:

D1: Xiaofan Lin, "Header and Footer Extraction by Page-Association", [online] 2002, pages 1-8, XP002533579, retrieved from the Internet: URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.6211> [retrieved on 2009-06-23].

As to auxiliary requests II and III, also filed with letter dated 29 September 2011, the Examining Division considered that they did not comply with Article 123(2) EPC.

Under the heading "Obiter Dictum", the Examining Division raised objections under Article 84 EPC and added some remarks relating to the lack of inventive step of the claimed subject-matter, when interpreted in the light of the description, and in particular concluded that the additional features recited in claim 1 of auxiliary requests II and III did not contribute to inventive step.

III. With the statement of grounds of appeal, the appellant filed new auxiliary requests I to IV and requested that

the decision under appeal be set aside and a patent be granted on the basis of the main request considered in the contested decision, or on the basis of one of the new auxiliary requests I to IV.

- IV. The appellant was summoned to oral proceedings to be held on 5 October 2016.
- V. In a communication pursuant to Article 15(1) RPBA, the Board addressed the appellant's requests and raised objections under Articles 84 and 56 EPC.
- VI. In reply to the Board's communication, the appellant filed with letter dated 21 September 2016 a new main request and new auxiliary requests I to IV, and, as a precautionary measure only, maintained the previous main request and auxiliary requests I to IV, renumbered and resubmitted as auxiliary requests V to IX.
- VII. At the oral proceedings, which were held as scheduled on 5 October 2016, the main request and auxiliary requests I to IV were admitted into the proceedings and the appellant withdrew auxiliary requests V to IX.
- VIII. Thus, the appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of the main request or of one of auxiliary requests I to IV as filed with letter dated 21 September 2016.
- IX. Claim 1 according to the main request reads as follows:
- "A method for identifying a header/footer of a document for facilitating structural legacy document conversion of the document, the method comprising:  
computing a textual variability comprising:

fragmenting content of the document into a number of text blocks;

identifying a relative position of each of the text blocks by allocating a relative vertical position per page of the document;

differentiating the text blocks at each relative position into different kinds of text blocks, wherein the different kinds of text blocks relate to differentiable text blocks; and

counting the number of text blocks of each relative position and counting the number of different kinds of text blocks at the respective relative position, wherein computing the textual variability of the content at the respective relative position comprises mathematically relating the number of different kinds of text blocks to the number of text blocks for computing a textual variability score (TVS);

wherein the method further comprises:

comparing the computed textual variability score with a predetermined textual variability indicative of a header/footer; and

associating contiguous contents having a computed textual variability score less than the predetermined textual variability wherein the associated contents are construed to comprise the header/footer."

Claims 2 to 4 are directly or indirectly dependent on claim 1.

Claim 5 reads as follows:

"An apparatus for identifying a header/footer in a document, the apparatus comprising:

a text fragmenter (14) for segregating the document into pages and differentiable text blocks on the pages, and wherein the text blocks are identifiable with respect to a relative zone within the page;

a processor (18) for computing a textual variability score (TVS) for the text blocks, and wherein the TVS comprises a relationship between a total number of text blocks and a number of different kinds of text blocks within a selected zone of the page; and

a selector (22) for selecting potential header/footer text content comprising a compilation of text blocks having a TVS indicative of header/footer content whereby the selected text content is set as the header/footer."

Claim 1 according to auxiliary request I differs from claim 1 of the main request in that the fragmenting step and the counting step read as follows:

"fragmenting content of the document into a number of text blocks arranged on pages of the document";

"counting the number of text blocks of each relative position and counting the number of different kinds of text blocks at the respective relative position, taking into account all the pages of the document, wherein computing the textual variability of the content at the respective relative position comprises mathematically

relating the number of different kinds of text blocks to the number of text blocks for computing a textual variability score (TVS)".

Claim 1 according to auxiliary request II differs from claim 1 according to the main request in that it ends with the following additional feature:

"wherein the mathematically relating comprises dividing the number of different kinds of text blocks by the number of text blocks".

Claim 1 according to auxiliary request III differs from claim 1 of auxiliary request II in that the counting reads as follows:

"counting the number of text blocks of each relative position and counting the number of different kinds of text blocks at the respective relative position taking into account the whole document, in particular, all the pages of the document, wherein computing the textual variability of the content at the respective relative position comprises mathematically relating the number of different kinds of text blocks to the number of text blocks for computing a textual variability score (TVS)".

Claim 1 according to auxiliary request IV differs from claim 1 of auxiliary request II in that the counting step reads as follows:

"counting the number of text blocks of each relative position and counting the number of different kinds of text blocks at the respective relative position, wherein computing the textual variability of the content at the respective relative position comprises



mathematically relating the number of different kinds of text blocks to the number of text blocks, taking into account the whole document, in particular, all the pages of the document, for computing a textual variability score (TVS);"

and in that the claim ends with the following additional feature:

"and wherein surrounding content of the associated contents is merged with the associated contents as header/footer content when a computed textual variability score for the merged surrounding and associated contents has a lower textual variability score than a computed textual variability score for the associated contents, wherein the surrounding content has a computed textual variability that is higher than the predetermined textual variability indicative of a header/footer."

The auxiliary requests also comprise apparatus claims 5 corresponding to the respective method claims 1.

X. The appellant's arguments relevant to this decision may be summarised as follows:

The claimed subject-matter related to a method for identifying the header/footer of a document for facilitating structural conversion of legacy documents. In structured documents, content was organised into delineated sections, such as document pages with suitable headers/footers. When converting unstructured documents to a structured format, headers/footers could generate an incorrect logical document and could also introduce "noise" which would affect further

processing, such as natural language processing (see paragraph 4 of the present application).

Although methods and apparatuses for identifying and extracting pagination constructs in the conversion of structured legacy documents had been known, the method according to claim 1 of the main request was fundamentally different from the approach disclosed in the prior art document D1. In particular, according to the method of claim 1, textual content at a specific relative position on a page of a document, i.e. a text block, was compared to the textual content at the same relative position on different document pages. For each position on a page, computation was then performed for associating the total number of text blocks at that position with the total number of different kinds of text blocks at the same position. The obtained textual variability score was then compared with a predetermined textual variability indicative of a header or footer. The approach described in the present application was computationally efficient and more robust than prior art methods.

Document D1 relied upon comparison of lines on neighbouring pages for identifying the particular relationship indicative of commonly configured headers/ footers and only disclosed comparing each line of a page with the same line on the previous and subsequent pages. The result of this comparison was a score indicating the similarity of lines located at the same position on neighbouring pages. In other words, document D1 only disclosed computing a score for each line. None of the steps recited in document D1 resulted in computing the number of text blocks at a relative

position and the number of different kinds of text blocks at the same position.

The algorithm used in document D1 to obtain a similarity score was relatively complex and thus computationally expensive. Moreover, the neighbouring page comparison technique according to document D1 could fail when the header/footer occurred very few times in a document. On the other hand, mathematically relating the number of different kinds of text blocks to the number of text blocks at a certain relative position on a page, as taught by the present application, was simpler and thus computationally more efficient. Furthermore, by taking into account the total number of text blocks, the method proved more robust even if the header/footer occurred very few times in a document.

When starting from document D1, the objective technical problem could thus be formulated as providing a more reliable method for identifying a header/footer of a document which facilitated structural conversion of legacy documents.

A reader trying to identify a header/footer might compare different pages in order to detect similar structures indicative of a header/footer. The same reader, however, would not count the number of text blocks located at the same relative position on the document pages and the number of different kinds of text blocks at the same position so as to establish a mathematical relationship between these numbers for computing a textual variability score. There was no evidence that mathematically relating these counts would be the most straightforward way of providing an alternative algorithm for identifying headers/footers.

In summary, a skilled person had no incentive to modify the teaching of document D1 so as to arrive at the claimed subject-matter in an obvious way. In particular, no document had been cited which would provide a hint prompting the skilled person to fragment the content into a number of text blocks, differentiate the text blocks into predetermined different kinds of text blocks and mathematically relate the number of different kinds of text blocks to the number of text blocks for computing a textual variability score. Consequently the subject-matter of claims 1 of all the requests involved an inventive step.

Further relevant arguments submitted by the appellant are referred to in the reasons for the decision below.

## **Reasons for the Decision**

1. The appeal is admissible.

### The invention

2. As explained, *inter alia*, in paragraph [0017] of the description of the original application, the purpose of the present invention is to detect a zone at the top or at the bottom of a page of a document, which corresponds to a pagination construct such as a header or footer zone. The detection of a header or footer zone is based on the assumption that in a header or footer zone "*the textual variety is much lower than in the body page of the document*".
  - 2.1 Paragraph [0017] identifies the following steps for header/footer detection: position allocation, text

normalisation, textual variability computation and header/footer zone detection.

- 2.1.1 A text fragmenter breaks a document, which has been converted, for instance into XML, into *"an order sequence of text fragments comprised of pages and text blocks. ... For a text document, each line suitably becomes a fragment ordered line-by-line"* (paragraph [0019]). As shown in Figures 3A and 3B, the line number or the y-coordinate provides the positional information of a text line or a block of text on a page. It is assumed that a header will likely occur in the top half of a page and a footer in the bottom half.
- 2.1.2 As explained in paragraph [0022], textual content is preferably normalised by replacing all digits [0-9] with a unique character.
- 2.1.3 Textual variability computation is explained in paragraph [0023]. The different positions of all the textual blocks in a document are listed. For each position, the number of text blocks occurring at a selected position and the total number of different blocks at the same position are computed, taking into account all the pages of the document. A textual variability score (TVS) is defined as the ratio between the number of different blocks and the total number of blocks at a particular position on the document pages.
- 2.1.4 After the TVSs for all the lines of the document have been determined, lines relating to potential header/footer data (i.e. lines with a low text variability score) are identified and further used to detect the header/footer zone. Once a first text element (line or block) likely to be part of the header/footer zone is identified on the basis of the TVS, the zone is

extended from this textual element by merging all contiguous content elements in order to find the largest possible header/footer zone (see paragraph [0028]).

- 2.1.5 As explained in paragraph [0029], the identification of a potential header/footer textual content element is accomplished by preselecting a given TVS threshold which has been empirically determined to give a likely identification of a header/footer content element.
- 2.2 Starting with the initial content element in the header and footer zones, a list with all contiguous potential textual content candidates is built. A new TVS for this augmented list is computed by dividing the total number of different types of text blocks in the list by the entire number of text blocks in the list (see paragraph [0032]).
- 2.3 Furthermore, as pointed out in paragraph [0034], surrounding line elements to the header/footer zone are investigated in order to verify whether line elements with higher variability, i.e. with variability higher than a threshold value, should be considered as components of the header/footer. Surrounding line elements that lower the overall TVS of the lines already identified as header or footer are included in the header or footer zone.

Main request

3. Claim 1 of the main request relates to a *"method for identifying a header/footer of a document for facilitating structural legacy document conversion of the document"*. The method comprises the following features:

- (a) computing a textual variability comprising:
  - (i) fragmenting content of the document into a number of text blocks;
  - (ii) identifying a relative position of each of the text blocks by allocating a relative vertical position per page of the document;
  - (iii) differentiating the text blocks at each relative position into different kinds of text blocks;
    - wherein the different kinds of text blocks relate to differentiable text blocks; and
  - (iv) counting the number of text blocks of each relative position and counting the number of different kinds of text blocks at the respective relative position,
  - (v) wherein computing the textual variability of the content at the respective relative position comprises mathematically relating the number of different kinds of text blocks to the number of text blocks for computing a textual variability score;
- (b) comparing the computed textual variability score with a predetermined textual variability indicative of a header/footer; and
- (c) associating contiguous contents having a computed textual variability score less than the predetermined textual variability wherein the

associated contents are construed to comprise the header/footer.

4. The amendments which distinguish the present claim 1 from claim 1 of the main request considered in the contested decision are essentially clarifications of features which also the Board had found unclear in its preliminary opinion. In particular, claim 1 now recites that the position of each of the text blocks is specified with respect to their vertical positions on a page. Furthermore, different kinds of text blocks are now defined as relating to differentiable text blocks.
  - 4.1 As to the interpretation to be given to the wording of claim 1 and in particular to the amended features, it is to be understood, as confirmed by the appellant, that text blocks, into which the content of a document is fragmented, can correspond to a text line or to a plurality of text lines. This interpretation is also supported by the fact that a text block is identified only by its vertical position. If a text line comprised more than one text block, it would be necessary to define also the horizontal position of a text block on a page.
  - 4.2 As to the feature that different kinds of text blocks relate to differentiable text blocks, the appellant has submitted that differentiable blocks were all blocks which were not identical. Thus, even minor differences between text blocks sufficed to define different kinds of text blocks.

*The contested decision*

5. In the contested decision, the Examining Division considered that automatically identifying headers and



footers involved a technical effect and that this was achieved by the method disclosed in document D1. However, the analysis of text to find textual variability, in the general sense of the claim, did not require any technical considerations since it was purely mathematical. In the Examining Division's view, technical considerations came into play only when technically implementing the non-technical features recited in the claim. However, technical considerations for implementing an algorithm of the kind shown in document D1 could not support an inventive step, because, apart from being trivial to the skilled person, they were identical to the ones involved in the implementation of D1.

6. In particular, the Examining Division developed two lines of argument leading to the conclusion that the subject-matter of claim 1 of the main request then on file lacked an inventive step.
  - 6.1 According to the first reasoning, the Examining Division considered that claim 1 covered also the possibility that the content of a text line was fragmented into words. In this case the two counts were meaningless for the determination of a header/footer and thus could not be related mathematically to provide a measure of textual variability. In other words, the claim as a whole did not have a technical effect since the features of the claim could not achieve the desired effect, i.e. identify a header/footer of a document.
    - 6.1.1 The above reasoning can no longer apply to the present claim 1 which no longer covers the fragmentation of a text line and necessarily implies that a text block cannot be smaller than a line (see point 4.1 above).

6.2 According to the second reasoning, a reader, paging through a document and wanting to know whether a certain text passage at, for instance, the bottom part of a page was a footer or belonged to the text body, would consider the variability of the text at this position, thereby implicitly comparing the number of different kinds of text blocks at this position with the total number of text blocks at the same position, without making any particular counts. For example, if a particular text block occurred identically at the same position multiple times on sequential pages, the variability being very low, then the reader would recognise the text as being a footer. Thus, the concept of using variability in the sense of comparing the number of different kinds of text blocks at a certain position with the total number of text blocks at the same position represented common knowledge.

6.2.1 Document D1 also disclosed an algorithm for determining textual variability in order to identify headers/ footers. As shown on page 5, table 2, certain errors still occurred in this algorithm. The skilled person, thus, had a clear incentive to modify the teaching of document D1.

6.2.2 In the Examining Division's opinion, the skilled person would just try out multiple ways of determining textual variability. The most straightforward way of doing so would be to bring the count of the total number of text blocks at a position and the count of the number of different kinds of text blocks at the same position into a mathematical relationship, as recited in claim 1.

7. The Board considers that automatically identifying headers and footers in an electronic document

constitutes a technical problem which essentially consists in identifying and possibly removing certain artefacts from the text body of the document. As explained in document D1, headers and footers may fragment the text flow in an electronic document that has been obtained from image or PDF files. It may also be useful to extract headers and footers from an electronic document so that they can be processed separately from the text.

- 7.1 Furthermore, both document D1 and the present application base the detection of headers and footers on the identification of certain features, such as their location on a page and the repetition of certain patterns, which are essentially dependent on the function of headers and footers in a document, and do not relate to their semantic content.
- 7.2 Document D1 deals with the problem addressed in the present application and in particular aims at detecting header/footer text lines.
- 7.2.1 The essential features of the teaching of document D1 can be summarized as follows:
- For example, the top five lines of a page are selected as the header candidates, while the bottom three lines are chosen as the footer candidates (D1, page 3, "Step 2").
  - Each candidate line will be quantitatively evaluated as to how well it qualifies as a header or footer (page 3, "Step 3").
  - The most stable feature of headers and footers is that they will repeat in neighbouring pages. The

problem is to quantitatively measure such repeats (see D1, page 4, section 2.2).

- Candidate lines, that is lines located at a given position on a given number of pages, are compared to determine their similarities (page 4, section 2.2).
- The similarity between two lines is calculated as the ratio between the number of matched characters and the larger of the numbers of characters in the two lines (page 4, section 2.2, expression (2)).

7.2.2 As it can be seen from the above summaries, both the present invention and document D1 make essentially the same assumptions as to what distinguishes a header or footer from the remaining content of a document. Document D1 compares a candidate line with lines at the same position on a limited number of neighbouring pages. The comparison is made on the basis of the matching characters present in two lines. On the other hand, the present invention assumes that the content of a text line in a header or footer is likely to be repeated throughout a document so that a limited number of "types" of content can be associated with a header or footer line.

7.2.3 As pointed out in the first paragraph of section 1 ("Introduction") of D1, headers and footers are common formatting elements in all kinds of documents. *"Besides reiterating key archival information such as author names, publication titles, page numbers, and release dates, they also serve decoration purpose by making the page layout more balanced and more visually appealing"*.

7.2.4 Although it seems easy for humans to locate headers and footers, it is technically challenging to build intelligent computer programs with similar capabilities. In fact, some documents have both footers and headers, some only have headers or footers, and some have neither of them. Furthermore, the headers/footers can contain the same text on all pages, such as journal or book titles, or various texts on different pages, such as page numbers and current article titles. (see page 2, paragraph 1, "Headers and footers exist in all kinds of formats").

7.2.5 Under the heading "2. Page-Association Based Header/footer Extraction", it is pointed out in D1 that, *"[a]lthough it is difficult to find stable page-level features that can be used to extract headers and footers, there does exist a relatively stable characteristic if we look beyond individual pages. Usually a document contains multiple pages, whose headers and footers are related to each other. The page-association based header/footer extraction is such an observation: Header/footers are text lines on the top/bottom of the pages with the same/similar counterparts in the neighboring pages. So instead of concentrating on individual pages, we inspect one page's relationship with its neighbors. In fact, this idea is in accordance with the way headers and footers are generated: The publishing or word-processing software allows the user to define rules to generate the headers and footers of continuous pages"* (underlining added).

Under the heading "2.2 Page-association based header/footer evaluation", it is reiterated in D1 that *"the most stable feature of headers and footers is that they will repeat in neighboring pages. The problem is how to*

*quantitatively measure such repeats".* Among the issues that have to be addressed, document D1 mentions that headers and footers can appear in numerous patterns, and that odd pages and even pages of the same document may have different headers.

7.2.6 As shown above, document D1 solves the problem of identifying a header or footer essentially by comparing text strings contained in two lines and calculating a "similarity score" which *"is high only if similar lines exist within the page window, no matter what pattern the header/footers are following"* (D1, page 4, lines 24 and 25).

7.3 In summary, document D1 points out that manual extraction of headers and footers from a document is time and labour consuming and that it is desirable to explore methods for the automatic extraction of headers and footers by page-association. According to document D1, such methods are essentially based on the realisation that the most stable feature of headers and footers is that they will repeat in neighbouring pages.

7.3.1 Starting from this everyday realisation that documents may have different headers/footers, but that an essential characteristic of headers/footers is that they are repeated in neighbouring pages, the skilled person wishing to develop a computer-implemented method for identifying a header/footer of a document will have first to find a parameter which reflects the repetitive nature of lines in a header/footer.

7.3.2 If a text line located at a certain vertical top or bottom position on a page is repeated throughout a document, it can be assumed that it belongs to a header or footer. A straightforward way to quantify the

repetitive nature of a text line is to find out how many times it occurs at that specific position in the document and how often its content changes throughout the document. If it never changes, there is just one kind of text line at that position for every page of the document. If it is different for odd and even pages, there are two kinds of text lines throughout the document. If the document is a book, a different header for each chapter may be expected.

7.3.3 From these general considerations, the skilled person will arrive at the trivial conclusion that the significant parameters for determining if a text line at a certain vertical position on a document page is repeated enough to be considered part of a header or footer are the number of times this text line occurs and how often it changes throughout the document. If the number of times a line at a certain position changes is small in comparison with its number of occurrences at the same position, it is plausible to assume that it belongs to a header/footer. If its number of occurrences is not significantly larger than the number of variations, it is unlikely to be part of a header/footer.

7.4 In summary, the textual variability score according to the claimed invention expresses nothing more than the number of times a text line changes in relation to the number of times it occurs in a document, and thus constitutes a straightforward embodiment of a parameter which reflects a generally known characteristic of a header/footer.

7.4.1 In the light of general knowledge common in the art and summarised in document D1, it would have been obvious to a skilled person, wishing to provide an alternative

solution to the problem of header and footer identification in a document, to arrive at a method falling within the terms of claim 1.

Hence, the subject-matter of claim 1 does not involve an inventive step within the meaning of Article 56 EPC.

7.4.2 The same applies to claim 5 which relates to an apparatus comprising features corresponding essentially to the claimed method steps.

#### Auxiliary requests I to III

8. Auxiliary request I differs from the main request only in that it is specified in feature (a)(i) that:

- the text blocks into which content of the document is fragmented are *"arranged on pages of the document"*, and
- the counting step (a)(iv) involves *"taking into account all the pages of the document"*.

8.1 Claim 1 according to auxiliary request II differs from claim 1 of the main request in that feature (a)(iv) is further specified as follows:

- *"wherein the mathematically relating comprises dividing the number of different kinds of text blocks by the number of text blocks"*.

8.2 Claim 1 according to auxiliary request III differs from claim 1 according to auxiliary request II in that:

- the counting step (a)(iv) involves *"taking into account the whole document, in particular, all the pages of the document"*.



8.3 The above features clarify certain aspects of the claimed method, such as the arrangement of the text blocks on the document pages, the actual mathematical expression for the textual variability and its computation on the basis of the whole document, which the Board has already taken into consideration when assessing the inventive step of claim 1 according to the main request.

Hence, the Board comes to the conclusion that none of auxiliary requests I to III complies with Article 56 EPC.

Auxiliary requests IV

9. The only substantial difference between the independent method claim 1 of auxiliary request IV and the ones of the higher ranking requests is that the former comprises the following additional step:

(d) wherein surrounding content of the associated content is merged with the associated contents as header/footer content when a computed textual variability score for the merged surrounding and associated contents has a lower textual variability score than a computed textual variability score for the associated contents,

(i) wherein the surrounding content has a computed textual variability that is higher than the predetermined textual variability indicative of a header/footer.

10. According to the appellant, auxiliary request IV was based on former auxiliary request IV filed with the statement of grounds of appeal. In addition, auxiliary

request IV defined the surrounding content as having a computed textual variability that was higher than the predetermined textual variability indicative of a header/footer. In the appellant's opinion, this amendment was supported by paragraph [0035] of the application as filed. Furthermore, the expression "compiled textual variability score" used in the previous auxiliary request IV had been replaced by the expression "computed textual variability score", as there was no difference between the two expressions. The latter amendment was supported by paragraph [0034] of the application as filed. As support for the former auxiliary request IV, the appellant had referred also to paragraphs [0034] and [0035] and to the paragraph bridging page 5 and 6 of the original application.

- 10.1 In the Board's opinion, it can be understood from the wording of claim 1 that the "*contents*" (i.e. "*associated contents*") referred to in feature (c) are text blocks identified by respective vertical positions on a page and located contiguously. As specified in the claim, each "*content*" has a textual variability calculated according to features (a)(iv).
- 10.1.1 The claim essentially indicates that text blocks which are contiguous with respect to their vertical positions are part of a header/footer if their respective textual variability scores are smaller than a predetermined threshold.
- 10.1.2 Claim 1 according to auxiliary request IV adds two further conditions for other "*contents*" which may be located on either side of the "*associated contents*", namely that this "*surrounding content*" has a textual variability score that is higher than the predetermined threshold (if this were not the case, it would be

associated content), and that the *"computed textual variability score for the merged surrounding and associated contents has a lower textual variability score than a computed textual variability score for the associated contents"*.

10.1.3 As to the latter condition, the Board considers that it could be interpreted in the sense that *"merged surrounding and associated contents"* constitute a new text block at a defined vertical position for which a textual variability could be computed as indicated in feature (a)(iv). On the other hand, the claim wording covers also the possibility that a textual variability score for the *"merged surrounding and associated contents"* is calculated on the basis of a number of different kinds of text blocks which is obtained by adding up the numbers of different kinds of text blocks counted at the relative vertical positions associated with said surrounding and associated contents.

10.2 Asked by the Board which interpretation of claim 1 was correct, the appellant submitted at the oral proceedings that the textual variability score of a merged surrounding and associated content should be computed by taking these contents as a new text block. In fact, in this case it would be possible to fulfil both conditions given in claim 1, namely that the textual variability score (TVS) of the surrounding content was to be above the predetermined threshold, while the TVS for the merged surrounding and associated content was to be lower than the TVS for the associated contents.

10.2.1 Although the appellant's interpretation of claim 1 appears plausible, in the Board's opinion, it finds no support in the original description.

10.2.2 According to paragraph [0032] "any contiguous textual content elements to the initial contact [sic - it should read content] element having a TVS lower than  $\theta$  also belong in the header/footer zone. Starting with the initial contact [sic] elements in the header and footer zones, a list with all the contiguous potential textual contact [sic] candidates is built. The augmented list comprises a new larger header/footer zone. A new TVS for the augmented list is computed 70 by applying the formula of Equation 1 for the entire number of text blocks in the list as a divisor for the total number of different types of text blocks in the list. (By construction, the new augmented list also has a variability score lower than  $\theta$ )" (underlining added). In other words, this passage of the description makes clear that the augmented list does not define a new text block.

10.2.3 Furthermore according to paragraph [0034], "surrounding line elements to the header/footer zone are further investigated for identifying previous and following elements whose variability is higher than  $\theta$ , but which line elements should also be considered nevertheless as components of the headers/footers. The merging 72 of potential surrounding lines in processor 24 is affected by determination that the preceding and following elements are appropriate for insertion into the header/footer zone if a lines insertion decreases the TVS of the new augmented list. More particularly, and with references to FIGURES 4A, 4B and 4C, it can be seen that the header zone augmented list is composed of line contact [sic] elements 108, 122 which have been merged as a result of their individual TVS. This surrounding element 96 is then next tested to determine if it can be inserted into the header zone and if the new TVS of

*the augmented list plus element 96 would be lower. Accordingly, the textual content, i.e., text blocks and number of different kinds of text blocks, is added into the list for computation of a new TVS. In this case element 96 does in fact lower the overall TVS of the augmented list plus surrounding line 96 so that the line content relative to 96 is in fact added to the header. The content of line element 143 is similarly tested, but the score of the augmented list plus surrounding line element 143 is 0.24 which is higher than the old score (0.23). Accordingly, the content element of the text occurring at the position 143 should not be added to the header zone and is not added to the augmented list. Line element 143 is not inserted and the merging of contiguous lines to identify the header zone then stops. The same procedure is effected for the initial contact [sic] elements, plus merge lines, plus surrounding lines for the footer zone" (underlining added).*

10.3 The above passages contain also some misleading information. In particular, the score (0.23) of the header list for the "line contact elements" 108 and 122 of Table 3 is not consistent with the corresponding numbers of blocks and of types of blocks given in Table 3, which give a score of 0.21. The same holds for the score (0.24) of the augmented list comprising the line element 143.

10.3.1 Furthermore, it is stated in paragraph [0034] that the element 96, having a TVS higher than 0, lowers the overall TVS of the augmented list plus surrounding line 96, although no value for the overall TVS is given. In fact, the overall TVS of the "line contact elements" 96, 108 and 122 and 96 calculated on the basis of the numbers given in Table 3 is 0.225, which is higher than

the TVS obtained for 108 and 122 (0.21). The appellant has failed to provide an explanation for these discrepancies.

- 10.4 In summary, the Board considers that claim 1 according to auxiliary request IV does not comply with Article 84 EPC because it is not clear from the wording of the claim how the textual variability score should be computed. Furthermore, the interpretation which appears to be more plausible and which is in conformity with the appellant's submissions has no support in the description. On the other hand, the interpretation that can be given to claim 1 in the light of the description does not confirm the numerical examples referred to above and thus does not seem to make any technical sense.
- 10.5 As claim 1 according to auxiliary request IV is unclear and finds no support in the application as originally filed, it does not comply with Article 84 EPC.
11. As none of the appellant's request can form the basis for the grant of a patent, the appeal has 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