

Internal distribution code:

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

**Datasheet for the decision
of 13 October 2015**

Case Number: T 0483/11 - 3.5.01

Application Number: 03786541.7

Publication Number: 1573562

IPC: G06F15/00, G06F17/00

Language of the proceedings: EN

Title of invention:

METHODS AND APPARATUS FOR SUMMARIZING DOCUMENT CONTENT FOR
MOBILE COMMUNICATION DEVICES

Applicant:

Arizan Corporation

Headword:

Document summary/ARIZAN CORPORATION

Relevant legal provisions:

EPC 1973 Art. 56

Keyword:

Decisions cited:

T 0258/03, T 0641/00

Catchword:

A feature does not automatically inherit the technical character of the context in which it occurs. The feature must, itself, make a contribution to the technical context or the technical aspects of the invention (technical inheritance fallacy - see points 2.7 and 2.8 of the reasons).



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 0483/11 - 3.5.01

D E C I S I O N
of Technical Board of Appeal 3.5.01
of 13 October 2015

Appellant: Arizan Corporation
(Applicant) 5775 Glenridge Drive,
suite 150 building C
Atlanta, GA 30328 (US)

Representative: Reichl, Wolfgang
MERH-IP
Matias Erny Reichl Hoffmann
Patentanwälte PartG mbB
Paul-Heyse-Strasse 29
80336 München (DE)

Decision under appeal: **Decision of the Examining Division of the
European Patent Office posted on 4 October 2010
refusing European patent application No.
03786541.7 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman W. Chandler
Members: P. Scriven
I. Beckedorf

Summary of Facts and Submissions

- I. This is an appeal against the Examining Division's decision to refuse European patent application EP03786541.7. The application concerns the generation of summary information for electronic documents.
- II. The following documents are referenced in this decision:
- WO 02/077855 A1 (D1)
EP-A-0 933 712 (D2)
WO 02/33584 A1 (D3)
- III. The Examining Division rejected the main and first and second auxiliary requests for lack of inventive step. It considered that the invention, as defined in claim 1 according to each of those requests, differed from D1 by the process of selecting summary information based on differences in paragraph size that was used when the document was both "unstructured" and "unformatted". This was held to be a non-technical user-requirement, the implementation of which would have been obvious to the skilled person. Moreover, this form of summarization was held to be an obvious extension of D1, particularly in view of the teachings of D3.
- IV. In the statement setting out the grounds of appeal, dated 11 February 2011, the appellant argued that D1 did not disclose the generation of a summary, let alone the summarization based on paragraph size used when the document was both unstructured and unformatted. Also, according to the appellant, the summarization was a technical process which did not depend on any user input.

- V. In a communication pursuant to Rule 15(1) RPBA, the Board gave its preliminary opinion that a document summary was not technical and expressed doubts as to whether the rules for the summary were anything more than mental acts. The Board also considered each of D1, D2, and D3 as potentially prejudicial to the patentability of the invention.
- VI. In reply to the Board's communication, the appellant filed, with a letter dated 14 September 2015, a set of amendments comprising a main request and first to third auxiliary requests.
- VII. Oral proceedings before the Board took place on 13 October 2015. The appellant's 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, or of one of first to third auxiliary requests, all filed with the letter dated 14 September 2015, the third auxiliary request corresponding to the second auxiliary request as rejected by the Examining Division.

For the course of the oral proceedings, reference is made to the minutes.

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

A method of generating summary information for an electronic document (400, 500, 600) for use by a mobile communication device (106), the method being performed by one or more servers of a network and comprising:

analyzing a content structure or properties within an electronic document (400, 500, 600), wherein the content structure comprises a table of contents, pages, slides, and/or worksheets;

generating document summary information which includes an assemblage of a plurality of summary entries comprising sections of the electronic document without sending the entire electronic document, wherein the summary entries are selected from the contents of the electronic document based on the analysis of the content structure or properties, wherein

if the electronic document has a predetermined content structure then using a structured document summarization process (300) for selecting the plurality of summary entries from the electronic document based on the predetermined content structure,

if the electronic document has no predetermined content structure but has differences in text formatting and/or paragraph formatting then using an unstructured document summarization process (302) for selecting the plurality of summary entries from the electronic document based on the differences in text formatting and/or paragraph formatting, and otherwise

if the electronic document has no predetermined content structure and no differences in text formatting and/or paragraph formatting then using an unformatted document summarization process (304) for selecting the plurality of summary entries from the electronic document based on differences in paragraph size; and providing the document summary information for a mobile communication device.

IX. Claim 1 of the first auxiliary request reads:

A method of generating summary information for an electronic document (400, 500, 600) for use by a mobile communication device (106), the method being performed by one or more servers of a network and comprising:

analyzing a content structure or content properties within an electronic document (400, 500, 600), wherein

the content structure is a table of contents, a plurality of document pages, a plurality of document slides, or a plurality of spreadsheet worksheets, and the content properties correspond to text formatting relating to font types, font sizes, font weights and font styles and to paragraph formatting relating to paragraph alignments and paragraph indents;

generating document summary information which includes an assemblage of a plurality of summary entries comprising sections of the electronic document without sending the entire electronic document, wherein the summary entries are selected from the contents of the electronic document based on the analysis of the content structure or content properties, wherein

if the electronic document has a predetermined content structure then using a first document summarization process (300) for selecting the plurality of summary entries from the electronic document based on the predetermined content structure,

if the electronic document has no predetermined content structure but has differences in text formatting and/or paragraph formatting then using a second document summarization process (302) for selecting the plurality of summary entries from the electronic document based on the differences in text formatting and/or paragraph formatting, and otherwise

if the electronic document has no predetermined content structure and no differences in text formatting and/or paragraph formatting then using a third document summarization process (304) for selecting the plurality of summary entries from the electronic document based on differences in paragraph size; and

providing the document summary information for a mobile communication device.

X. Claim 1 of the second auxiliary request reads:

A method of generating summary information for an electronic document (400, 500, 600) for use by a mobile communication device (106), the method being performed by one or more servers of a network and comprising:

- receiving a request from the mobile communication device (106) for the electronic document (400, 500, 600);

- analyzing the electronic document (400, 500, 600) on the basis of a Document Object Model (DOM) associated with the electronic document (400, 500, 600), said Document Object Model (DOM) being a uniform representation of the content of the electronic document (400, 500, 600) in a hierarchical structure which allows for extraction of a particular part of the electronic document;

- generating document summary information which includes an assemblage of a plurality of summary entries comprising sections of the electronic document, wherein the summary entries are selected from the contents of the electronic document based on the analysis of the electronic document the basis of the Document Object Model (DOM); and

- transmitting the document summary information to the mobile communication device (106) without sending the entire electronic document;

wherein, if the electronic document includes a table of contents, then generating the document summary information includes selecting the plurality of summary entries from the electronic document based on the entries of the table of contents,

wherein, if the electronic document does not have a table of contents but has a content structure being a plurality of document pages or a plurality of spreadsheet worksheets, then generating the document

summary information includes selecting the plurality of summary entries from the electronic document based on the content structure of the electronic document such that the document summary information includes one entry for each spreadsheet worksheet of the plurality of the plurality of spreadsheet worksheets or the summary entries each correspond to respective contiguous ranges of pages of the plurality of pages of the electronic document,

wherein, if the electronic document does not have a table of contents and has no content structure being a plurality of document pages or a plurality of spreadsheet worksheets but has differences in text formatting relating to differences in font types, font sizes, font weights and font styles and/or differences in paragraph formatting relating to differences in paragraph alignments or paragraph indents, then generating the document summary information includes selecting the plurality of summary entries from the electronic document based on differences in the text formatting and/or differences in the paragraph formatting, and otherwise

wherein, if the electronic document does not have a table of contents and neither has a content structure being a plurality of document pages or a plurality of spreadsheet worksheets nor has differences in text formatting relating to differences in font types, font sizes, font weights and font styles and/or differences in paragraph formatting relating to differences in paragraph alignments or paragraph indents, then generating the document summary information includes selecting the plurality of summary entries from the electronic document based on differences in paragraph size.

XI. Claim 1 of the third auxiliary request reads:

A method of generating summary information for an electronic document (400, 500, 600) for use by a mobile communication device (106), comprising:

receiving a request from the mobile communication device (106) for the electronic document (400, 500, 600),

analyzing the electronic document (400, 500, 600) on the basis of a Document Object Model (DOM) associated with the electronic document (400, 500, 600), said Document Object Model (DOM) being a uniform representation of the content of the electronic document (400, 500, 600) in a hierarchical structure which allows for extraction of a particular part of the electronic document,

generating document summary information which includes an assemblage of a plurality of summary entries selected from the contents of the electronic document based on the analysis of the Document Object Model, and

transmitting the document summary information to the mobile communication device (106), characterised in that

the Document Object Model (DOM) is analyzed by performing an unformatted document summarization process (304) including selecting the plurality of summary entries from the electronic document based on an examination of paragraph sizes and paragraph text patterns of the electronic document and determining which paragraphs are most likely to be section identifiers that can be used as summary entries for the document summary information, wherein shorter paragraphs are identified with a higher likelihood to be section identifiers than longer paragraphs, said unformatted document summarization process (304)

comprising:

traversing the Document Object Model (DOM) associated with the electronic document to collect paragraph information from the electronic document, the paragraph information including the number of characters in the paragraph and the text contained in the paragraph,

determining whether the electronic document is summarizable or unsummarizable on the basis of the variance in the size of the paragraphs in the electronic document, wherein the electronic document is determined to be unsummarizable if it is determined that there is not sufficient variance in the size of the paragraphs in the electronic document on the basis of the ratio of the size of the largest paragraph to the size of the smallest paragraph, the document being determined to be unsummarizable if the ratio of the size of the largest paragraph to the size of the smallest paragraph is not greater than a configured minimum threshold and the document being determined to be summarizable otherwise,

organizing the paragraph information into groups of information for paragraphs that have the same number of characters, each parameter information group comprising a paragraph size and text from the paragraphs in the electronic document that contain a number of characters equal to the paragraph size,

ordering the paragraph information groups into an ascending order according to the paragraph sizes, wherein the paragraph information groups that specify smaller paragraph sizes have higher orders so that the paragraph information groups are ordered such that text from the paragraphs that are most likely to be section identifiers is contained in the groups of the highest orders, while text from the paragraphs that are least likely to be section identifiers is contained in the

*groups of the lowest order, and
generating the document summary information for the
electronic document using the paragraph information
group having the highest order.*

XII. The appellant's arguments in the oral proceedings can be summarized as follows:

The invention in claim 1 of the main request differed from D1 by:

(i) the unstructured summarization based on differences in paragraph formatting;

(ii) the unformatted summarization based on differences in paragraph size.

Paragraph formatting in claim 1 meant paragraph alignment or indentation, which was different from any text formatting (for example font size, font type, or font style) applied to a paragraph.

The summarization in D1 did not support unformatted documents, such as plain text documents (.txt files). If an unformatted document were submitted to the server of D1, the result would be an error code (figure 8, steps 154 and 156). By contrast, the method of claim 1 would produce a summary even for unformatted documents. Thus, the invention extended the types of documents that could be summarized. This was a technical effect.

The method of generating a summary was used in a mobile communication system. The summary information was generated and transmitted by a server in response to a client's request for the whole document. By sending only a summary and not the whole document, the data to be transmitted was reduced. Furthermore, the processing

burden on the mobile device was reduced. In other words, the summarization according to the invention was part of a technical context, and provided the technical effect of allowing the user of a mobile device to access documents that would otherwise have been too large.

Reasons for the Decision

1. *The invention*

1.1 At the date of the invention, mobile data connections were slow and mobile devices had limited processing and display capabilities. At the same time, electronic documents (word processor, worksheet, and spreadsheet documents) were large and contained "rich" content. Therefore, there was a need for a smaller, summary version of electronic documents for use by mobile communication devices.

1.2 According to the invention, the summary is generated by a server (figure 1, reference numeral 100) in response to a request from a mobile communication device (106) and is transmitted to the mobile device (page 4, lines 15 to 20). The user of the mobile device can use the summary to navigate the electronic document and request content corresponding to the summary entries from the server. This precludes the need to send the entire document to the mobile device (page 4, line 21 to page 5, line 3), at least initially.

1.3 The server generates the summary by selecting content from the electronic document. It does this using one of

three processes (figure 3).

If the document has what the application calls "content structure", e.g. high-level descriptive information such as a table of contents, this information is used as a summary (page 5, lines 9 to 18). That is the "structured document summarization process".

If the document does not contain such information, any text formatting or paragraph formatting is analysed in order to find "section identifiers" (headers and titles) in the document which are used as summary entries. This "unstructured document summarization process" operates on the assumption that section identifiers are formatted differently from the text body, e.g. using a larger font size (page 6, line 28 to page 7, line 7).

If the document contains neither "predetermined content structure" nor text or paragraph formatting information, or if all the text is formatted identically, the "unformatted document summarization process" is used. This operates on the basis of differences in paragraph size: shorter paragraphs (those having few characters) are more likely to be section identifiers than longer paragraphs (page 7, lines 23 to 30).

2. *Main request, inventive step*

2.1 The Board considers D1 to be an appropriate starting point for assessing the inventive step of claim 1 according to the main request. D1 has, like the invention, the aim of making the content of an electronic document accessible on a mobile device (page

1, line 18 to page 2, line 7). In D1, the client (12) submits a document to the server (10) for processing. The server processes the document to extract sections of document content and a table of contents (page 8, lines 3 to 7; page 13, lines 18 and 19). If there is no table of contents available, the server creates one by analysing the document. The content excerpts and table of contents are transmitted to the mobile device (12), on request (figure 12; page 27, lines 5 to 7; page 30, lines 18 and 19; claim 7). The table of contents can be used as navigational information to request the content (claim 8).

- 2.2 In the statement of grounds, the appellant argued that D1 did not disclose the generation of a summary, but at oral proceedings before the Board it was common ground that the table of contents (TOC) in D1 represented a summary in the sense of the invention as claimed and described in the application. It was also common ground that D1 disclosed the structured summarization and the unstructured summarization based on differences in text formatting, as defined in claim 1 of the main request.

At oral proceedings, the appellant identified the following differences of the invention over D1:

(i) "unstructured document summarization" based on differences in paragraph formatting;

(ii) "unformatted document summarization" based on differences in paragraph size.

- 2.3 The Board notes that summarization based on differences in paragraph formatting is defined in claim 1 alongside an alternative to summarization based on text formatting. It is enough for any one of these to belong

to the prior art. Furthermore, the Board does not agree with the appellant that "paragraph formatting" is limited to paragraph alignment or indentation, but considers that this broad term covers text formatting for whole paragraphs in D1 (page 20, lines 14 and 15; figure 15). That notwithstanding, the Board sees the TAB character in D1 (page 20, line 25 to page 21 line 3) as a disclosure of paragraph indentation. The TAB character in D1 is used as a section identifier when constructing a table of contents in the unstructured summarization process.

2.4 Thus, the Board agrees with the Examining Division that the only difference between the method of claim 1 and D1 is the "unformatted document summarization process" for extracting summary information based on differences in paragraph size.

2.5 The appellant argued that the document summarization in claim 1 was technical since it was part of a technical context, namely a mobile communication system. Moreover, the summarization was provided in order to overcome the technical limitations of such a system. The "unformatted document summarization process", in particular, was technical for those same reasons. It allowed a larger class of documents to be summarized and used in the context of the mobile communication system.

2.6 The Board does not dispute that the claimed method appears in a technical context. The method is performed by technical means (one or more servers of a network), and, therefore, has technical character. This is relevant to the question of whether the invention is an invention in the sense of Article 52(1) EPC (T 258/03

"Auction method/HITACHI", OJ EPO 2004, 575).

However, the question of inventive step, requires an assessment of whether the invention makes a technical contribution over the prior art. Features which do not make such a contribution cannot support the presence of an inventive step (T 641/00 "Two identities/COMVIK", Headnote I, OJ EPO 2003, 352).

- 2.7 In the present case, the contribution of the invention does not lie in the use of document summarization in a mobile communication system. That is already in the prior art. The contribution lies rather in the algorithm for extracting summary information from the electronic document, more specifically in the manner in which section identifiers are assumed in a text that has no differences in formatting. In the Board's view, this is not technical. It is a mental act, such as would be performed by a human when reading a text.
- 2.8 Put in the technical context of the mobile communication system, the unformatted document summarization has the consequence that a larger class of documents can be summarized. However, the Board does not consider this to be a technical effect. The Board does not share the appellant's view that a feature automatically inherits the technical character of the context in which it occurs. The feature must, itself, make a contribution to the technical context, or the technical aspects of the invention.
- 2.9 For these reasons, the Board takes the view that the "unformatted document summarization process" does not make a technical contribution over D1. Furthermore, the Board considers that the implementation of this

functionality would have been straightforward, using routine programming methods.

2.10 Therefore, the Board concludes that the invention as defined in claim 1 according to the main request lacks inventive step (Article 56 EPC).

3. *First auxiliary request, inventive step*

3.1 At oral proceedings, the appellant explained that the amendments to the first auxiliary request were meant to clarify some of the terms used in the main request. The distinguishing feature of the invention according to claim 1 was, as in the main request, the "unformatted document summarization process". Therefore, the Board considers that the same reasons apply to this request, and reaches the same conclusion as to inventive step.

4. *Second auxiliary request*

4.1 In claim 1 of the second auxiliary request, the summary entries are selected from the contents of the electronic document based on an analysis of a Document Object Model (DOM). This is disclosed in D1, see page 12, line 7 to page 13, line 2.

Claim 1 of the second auxiliary request also sets out the steps of receiving a request from the mobile communication device for the electronic document and transmitting the document summary information to the mobile communication device without sending the entire electronic document. This is already disclosed in D1 (see point 2.1 above).

At oral proceedings before the Board, the appellant submitted that claim 1 of the second auxiliary request differed from D1 by the "unformatted document summarization process", and the Board agrees. As concluded with regard to the main request, this feature does not provide an inventive step.

5. *Third auxiliary request*

- 5.1 It is common ground that the invention defined by this version of claim 1 differs from D1 by the "unformatted document summarization process". However, this process is defined in more detail than in the higher-ranked requests.

According to claim 1, it is determined whether the document is summarizable or unsummarizable on the basis of the "variance" in the size of the paragraphs in the electronic document. If the ratio of the size of the largest paragraph to the size of the smallest paragraph is not greater than a threshold, the document is determined to be unsummarizable.

If the document is summarizable, the paragraphs are organized into groups, so that each group contains paragraphs having the same number of characters. The groups are ordered according to paragraph size, so that paragraphs that are most likely to be section identifiers (i.e. short paragraphs) are contained in groups of the highest order.

The summary information is generated using the paragraphs in the group of the highest order.

5.2 In the Board's view, the above features relate to the algorithm of extracting information from an unformatted document. The Board does not see that they produce a technical effect, and, therefore, they do not contribute to inventive step.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chairman:



T. Buschek

W. Chandler

Decision electronically authenticated