

mail, Microsoft® Word documents, Microsoft® Powerpoint documents, news group postings, multimedia objects, Graphics Interchange Format images and/or Shockwave Flash files. The presentation semantics may be implicit or explicitly specified in a language known in the art or future developed such as, but not limited to, CSS. The presentation semantics for a document if explicitly specified may be included with in the same file along with the data or may be specified in external file(s) or a combination of both. An external presentation semantics file for a document may be herein referred to as presentation file. Thus, while our disclosure and the examples of use given herein are sometimes described in terms of HTML and CSS, this is to be understood to be an example of the use and utility of the disclosures, and is not intended to imply any limitation in the scope of their use. To the contrary, the disclosure here disclosed should be understood to be applicable as well to any document, file or combination of files assembled into a unit known in the art or future developed containing text, audio and video objects, images and other multimedia objects with optional presentation semantics instructions.

[0017] The present disclosure generally relates to methods and systems for searching data sources for information. More particularly, it relates to methods for vision-based identifying, extracting and indexing segments or parts of a document matching paragraph, table, list, menu, fixed width text, key/value, graph/chart, question/answer, timeline, and interactive (hereinafter referred to as “visual elements”) types similar to the way human beings do by looking at the document on the display interface.

[0018] In one ~~aspect~~part of the disclosure, a human being identifies and/or extracts visual elements in a document of visual element types from the rendering of the document on a display interface such as a monitor or on paper. In ~~still~~ another ~~aspect~~part of the disclosure the system automatically identifies and/or extracts visual elements in a document of visual element types with the aid of blocks and profiles. In yet another ~~aspect~~part of the ~~invention~~disclosure, a combination of human being and system may be used in the identification and extraction of visual elements.

[0019] A block is a logical unit of a document. A profile is a set of rules designed in such a way that it identifies and classifies a matching block(s) into a visual element type

exactly like a human being would classify the block(s) by visually looking at the block(s) when it is displayed as part of the document on the display interface.

[0020] Disclosed embodiments seek~~A further object of the present disclosure is to index and rank found visual elements.~~

[0021] Disclosed embodiments seek~~A further object of the present disclosure is to provide interfaces to the user to search for information relating to a topic by restricting the results to one or more visual element types. In this, the search interface provides suggestions to the user during the search query inputting phase and further more suggestions on the results report.~~

[0022] A further disclosed embodiment~~object of the present disclosure is to displays~~ prioritized search results in response to user search query, as horizontal list and/or vertical list and/or in a grid.

[0023] A further disclosed embodiment~~object of the present disclosure is to displays~~ a brief summary as paragraph text for each search result. The summary is that portion or portions of the text in the document that contain the keywords from the search query. In ~~another aspect~~one part of the present disclosure, the summary results are shown in the same visual type as found in the original document i.e. summary of a table visual type result in a table as found in the original document etc., In ~~still yet another aspect~~part, the result summary is shown in the same visual type and presentation semantics as found in the original document.

[0024] A further intention~~object~~ of the present disclosure is to show advertisements on the results report when a search is performed by the user using a search query which further includes visual element types. In yet another ~~aspect~~part of the present disclosure, marketers can bid for advertisement space on the results report by search query which further includes visual element type(s). In still yet another ~~aspect~~part of the present disclosure, document author or document owners may pay a fee for the document visual elements to be indexed and included as part of search results.

[0024a] According to one aspect, there is provided method for processing documents, the method comprising: generating preliminary metadata for a document, the document having source content and presentation semantics; dividing the document into one or more identified blocks using the preliminary metadata; preparing a block list of the one or more identified blocks; processing the one or more identified blocks in the block list using block operations; and identifying at least one title block item for at least one of the identified blocks in the block list.

[0024b] According to another aspect, there is provided a method for processing and identifying documents in accordance with their relevance to a search query, the method comprising: generating preliminary metadata for a document, the document having source content and presentation semantics; dividing the document into blocks using the preliminary metadata; preparing a list of the identified blocks; processing the blocks in the block list using block operations; identifying a title block for each block in the block list; identifying at least one inline visual element within a block item in the block list using predefined data and metadata rules, if an inline visual element exists; identifying at least one block visual element within at least one of the blocks in the block list using predefined profiles, if a block visual element exists; generating an index of the found visual elements; receiving search query data via a Graphical User Interface (GUI) wherein the search query data comprises at least one selected visual element type; and generating a response to the search query with identification of each document that is determined to be relevant to the search query.

DESCRIPTION OF THE DRAWING FIGURES

[0025] FIG. 1 is a block diagram illustrating an exemplary search engine system in accordance with an embodiment of the present disclosure.

[0026] FIG. 2 is a block diagram illustrating an exemplary search engine computing device of FIG. 1.

[0027] FIG. 3 depicts exemplary document source content data.