



- (51) **International Patent Classification:**  
G06F 17/20 (2006.01)
- (21) **International Application Number:**  
PCT/US2012/059917
- (22) **International Filing Date:**  
12 October 2012 (12.10.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US).
- (72) **Inventors:** SIMSKE, Steven J.; 3404 E. Harmony Rd., Fort Collins, Colorado 80528-9544 (US). STURGILL, Malgorzata M.; 3404 E. Harmony Rd., Fort Collins, Colorado 80528-9544 (US).
- (74) **Agents:** BURROWS, Sarah et al.; Hewlett-Packard Company, Intellectual Property Administration, 3404 E. Harmony Road, Mail Stop 35, Fort Collins, Colorado 80528-9599 (US).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

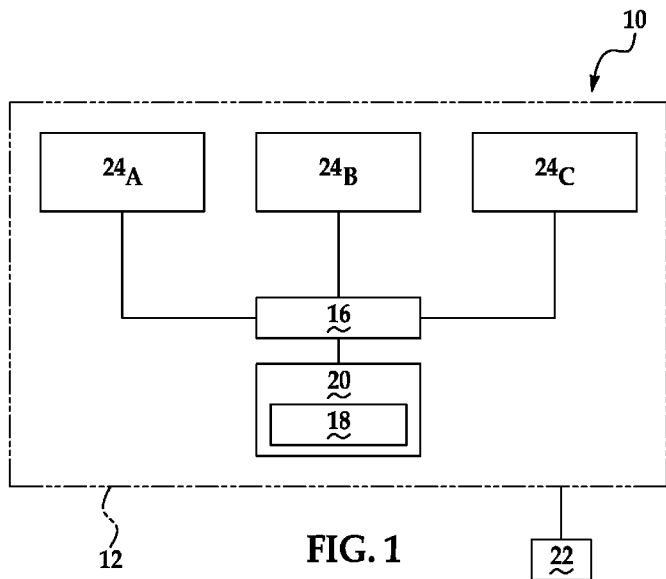
**Declarations under Rule 4.17:**

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

**Published:**

- with international search report (Art. 21(3))

(54) **Title:** A COMBINATORIAL SUMMARIZER



**FIG. 1**

(57) **Abstract:** A combinatorial summarizer includes a plurality of summarization engines, a processor in selective communication with each summarization engine, and computer readable instructions executable by the processor and embodied on a tangible, non-transitory computer readable medium. Each summarization engine is to select a respective plurality of sentences, and generate a relative rank and an associated weight for each sentence of the respective plurality of sentences. The computer readable instructions include instructions to determine a combined weight for each sentence of each respective plurality of sentences. The combined weight is based upon the respective associated weight and a respective relative human rank for each sentence in a set of sentences including all sentences of each respective plurality of sentences. The computer readable instructions further include instructions to determine a total weight for each summarization engine based, respectively, upon the combined weights for each sentence of each respective plurality of sentences.

WO 2014/058433 A1

## A COMBINATORIAL SUMMARIZER

### BACKGROUND

[0001] Summarizers are computer-based applications that analyze documents (e.g., text, html, email, etc.), extract main ideas or sentences from the analyzed documents, and arrange the extracted main ideas or sentences into a summary. Some summarizers enable the summary to be theme-oriented, structure-oriented, or concept-oriented. Different summarizers may provide a different summary for the same document.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Features and advantages of examples of the present disclosure will become apparent by reference to the following detailed description and drawings, in which like reference numerals correspond to similar, though perhaps not identical, components. For the sake of brevity, reference numerals or features having a previously described function may or may not be described in connection with other drawings in which they appear.

[0003] Fig. 1 is a schematic depiction of an example of a combinatorial summarizer; and

[0004] Fig. 2 is a schematic depiction of an example of the combinatorial summarizer implemented as part of a cloud computing system.

### DETAILED DESCRIPTION

[0005] The present disclosure relates generally to combinatorial summarizers.

[0006] When different summarizers provide different summaries, human evaluation of the different summaries may be used to rank the individual summaries. This approach provides a binary, qualitative datum that limits the outcome of the evaluation to the selection of one summarizer over another. In the examples disclosed herein, however, a combinatorial approach is utilized, which combines quantitative method(s) for evaluating the summarizers with the

advantages of qualitative human feedback. The combinatorial approaches disclosed herein enable the summarizers and their respective outputs to be combined together for significantly improved performance, accuracy, and summary robustness.

[0007] Examples of the combinatorial summarizer disclosed herein utilize the previously mentioned combinatorial method(s) to evaluate two or more summarization engines. The individual summarization engines analyze a document, and provide a summary output to a processor operatively connected thereto. In some examples, the summary outputs may be uniform in that each summarization engine is programmed to rank the same number of sentences from the document, but may be different in that the ranked sentences may vary from one summarization engine to another. In other examples, the summarization engines may be programmed to rank different numbers of sentences, and the processor may be programmed to crop the number of sentences ranked by each summarizer to any desirable number. The processor, executing computer readable instructions, utilizes these output summaries in conjunction with relative human ranking data for the same document in order to evaluate the summarization engines. The combinatorial summarizer and the combinatorial method disclosed herein provide a non-biased and quantitative approach for evaluating the summarization engines. In addition, some examples of the combinatorial method apply meta-algorithmic patterns, which combine at least some of the information from the various summary outputs in order to evaluate various combinations of the summarization engines.

[0008] Referring now to Figs. 1 and 2, examples of the combinatorial summarizer 10, 10' are shown. The combinatorial summarizers 10, 10' each include a combination of hardware and programming in order to perform examples of the combinatorial summarization methods disclosed herein. The hardware and programming may be part of a local computing system (e.g., reference numeral 12 in Fig. 1) or part of a cloud computing system (e.g., reference numeral 14 in Fig. 2), each of which will be discussed in further detail herein. In other instances, the

hardware and programming may be part of a distributed computing system, a virtualized computing system, or a peer-to-peer computing system.

[0009] Hardware generally refers to processors (e.g., processor 16), servers, and other computer hardware systems. The processor 16 is capable of executing programming for performing steps of the combinatorial methods disclosed herein.

[0010] Programming generally refers to computer readable instructions 18 embodied on a non-transitory, tangible computer readable medium 20 that, when executed, carry out a specific function. Examples of the instructions 18 disclosed herein may be realized in any non-transitory, tangible computer readable media 16 for use by or in connection with an instruction execution system (e.g., computing systems 12, 14), such as a computer/processor based system, or an ASIC (Application Specific Integrated Circuit), or another system that can fetch or obtain the logic from computer readable media 20 and execute the instructions 18 contained therein.

[0011] The non-transitory, tangible computer readable media 20 may be any media that is capable of containing, storing, or maintaining programs and data for use by or in connection with the computing systems 12, 14. It is to be understood that the media 20 may be integrated in the same device as the processor 16, or it may be separate from, but accessible to the respective computing system 12, 14 and processor 16. Examples of computer readable media 20 may include any one of many physical media such as, for example, electronic, magnetic, optical, electromagnetic, or semiconductor media. More specific examples of suitable computer readable media 20 include a portable magnetic computer diskette such as floppy diskettes or hard drives, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM), or a portable CD, DVD, or flash drive.

[0012] Still further, the computer readable instructions 18 may be part of an installation package that can be executed by the processor 16 to implement at least some steps of the combinatorial summarization methods disclosed herein. In these instances, the medium 20 may be the previously mentioned portable

medium, such as a compact disc (CD), a digital video disc (DVD), or a flash drive; or the medium 16 may be a memory maintained by a server from which the installation package can be downloaded and installed on the computing system 12. In another example, the computer readable instructions 18 may be part of an application or applications already installed on the computing system 12. In this other example, the medium 20 may include integrated memory, such as the previously mentioned hard drive.

[0013] Referring now specifically to Fig. 1, the example of the combinatorial summarizer 10 is shown as a local computing system 12, which may be a standalone computing system and/or part of a network of computing systems. The local computing system 12 may be any personal computer, portable computer, content server, a network PC, a personal digital assistant (PDA), a smart cellular telephone or any other computing device that is capable of performing the functions of the combinatorial summarizer 10 disclosed herein. The local computing system 12 is operatively connected to an input device 22, which may be a keyboard or a keypad, mouse, touchscreen, etc. which enables a user to input information into the computing system 12.

[0014] As shown in Fig. 1, the combinatorial summarizer 10 includes a plurality of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. While three summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> are shown in Fig. 1, it is to be understood that any number of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be part of the combinatorial summarizer 10. In the example shown in Fig. 1, each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is a separate computer program (i.e., set of computer readable instructions) that is executable by the processor 16. In this example, the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be stored in the computing system 12 on any of the examples of the computer readable medium 20 discussed herein. In another example that is not shown, one or more of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be physically separate from the computing system 12, but in selective communication with the processor 16, which receives data (i.e., summary outputs) from the respective summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In this example, the summarization engines

24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be on separate systems 12. In still another example, the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be part of a distributed, virtualized, or cloud computing system that the processor 16 of the local computing system 12 can operatively communicate with. For example, one summarization engine 24<sub>A</sub> could be resident on a mobile device and connected to the processor 16 while the other summarization engines 24<sub>B</sub> and 24<sub>C</sub> are resident on other devices that are automatically or volitionally connected to the processor 16.

[0015] The execution of any of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> results in the analysis and summarization of a user-selected document (e.g., text, html, email, etc.). In the examples disclosed herein, each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> summarizes the document and generates a respective summary output. Generating the summary output involves selecting a set number of sentences from the document and ranking each of the selected sentences from the most salient sentence to the least salient sentence. The rankings may vary from one summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> to the next depending, at least in part, on the programming. The number of sentences in the sets respectively selected by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be the same or different. In some examples, the processor 16 running computer readable instructions 18 may be programmed to crop the number of sentences from one or more summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> so that the total number of sentences provided by all of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> does not exceed the number of sentences in the original document.

[0016] The summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> also associate each of the relative ranks with a weight. The weighting scheme may assign weights in inverse order to the rank, where the top ranked, or most salient sentence (e.g., sentence number 1 of 10) receives the highest weight, and the bottom ranked, or least salient sentence (e.g., sentence number 10 of 10) receives the lowest weight. Other weighting schemes may also be utilized. As an example, the 15 sentences selected as the most significant sentences may be ranked on a scale of 1-15, with 1 being the least significant sentence and 15 being the most significant sentence.

In this example, the weighting scheme may multiple the ranks by a constant weight factor (e.g., a predetermined number, such as 10) so that the most significant sentence is associated with the highest weight, and the least significant sentence is associated with the lowest weight. In another example, the weights could be based on interval probability measures. In still another example, the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> could provide their own confidence values, although the processor 16 running computer readable instructions 18 may be programmed to normalize these weights so the weighting is the same for each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. When the processor 16 crops the number of sentences selected by any particular summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, new relative and absolute weights of the sentences that remain may be obtained.

[0017] It is to be understood that the set number of sentences to be selected and the weighting scheme are the same for each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. As such, the summary outputs from the respective summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> are somewhat uniform. However, it is to be understood that the summary outputs may not be identical, at least in part because each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may select different sentences from the document as being the most salient sentences.

[0018] In an example, each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> analyzes the text of an article containing 35 sentences. Each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> includes computer readable instructions to select what it deems to be the 10 most significant sentences of the 35 sentences, computer readable instructions to generate a relative rank for each of the selected sentences (where the rank of 1 being the most significant and the rank of 10 being the least significant), and computer readable instructions to associate a particular weight with the ranked sentences (where the weight is the inverse of the rank). Examples of summary outputs from the three summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> for this example are shown below in Tables 1-3.

**Table 1: Summary Output for Summarization Engine 24<sub>A</sub>**

<u>Rank</u>	<u>Sentence Number in the Document</u>	<u>Weight</u>
1	4	10.0
2	7	9.0
3	3	8.0
4	14	7.0
5	25	6.0
6	9	5.0
7	1	4.0
8	33	3.0
9	19	2.0
10	35	1.0

**Table 2: Summary Output for Summarization Engine 24<sub>B</sub>**

<u>Rank</u>	<u>Sentence Number in the Document</u>	<u>Weight</u>
1	7	10.0
2	1	9.0
3	6	8.0
4	29	7.0
5	19	6.0
6	4	5.0
7	5	4.0
8	3	3.0
9	33	2.0
10	14	1.0

**Table 3: Summary Output for Summarization Engine 24<sub>C</sub>**

<u>Rank</u>	<u>Sentence Number in the Document</u>	<u>Weight</u>
1	1	10.0
2	14	9.0
3	4	8.0
4	3	7.0
5	7	6.0
6	19	5.0
7	25	4.0
8	30	3.0
9	9	2.0
10	24	1.0

[0019] The weights in each of Tables 1-3 sum to 55. Dividing by 55 normalizes the weights to 1.0. It may be desirable to normalize the weights in a similar manner when the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> report different numbers of sentences, different weighting schemes, or when some of their sentences are cropped or dropped as more summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> are added to the overall set.

[0020] The processor 16 receives the output summary (including the relative rank and the associated weight) from each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. The processor 16, running computer readable instructions 18, is capable of generating a set of sentences based upon the output summaries. In particular, the processor 16 compiles a list of all of the sentences respectively ranked by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. The set of sentences that is compiled is a complete list of the sentences that were ranked by any of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, whether a sentence was ranked by one of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> or more than one of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. For the previous example, the set of sentences would include 15 sentences, namely sentence numbers 1, 3, 4, 5, 6, 7, 9, 14, 19, 24, 25, 29, 30, 33, and 35. The number of sentences included in the set of sentences may vary depending, at least in part, on the number of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, the set number of sentences to be selected by each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, and the number of sentences that are ranked by multiple summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In some instances, the number of sentences provided by each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be cropped/reduced until the number of unique sentences provided by the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (i.e., the number of sentences in the set) is below some desirable number.

[0021] The processor 16 is also capable of receiving human evaluation data from one or more human evaluators using the input device 22. The human evaluator(s) retrieve or are provided with the original document (or a copy thereof) and the set of sentences (e.g., the list of all of the sentences ranked by the

summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>). The human evaluator may retrieve the document and set of sentences from the local computing system 12. The human evaluator may also be provided with the document and the set of sentences. Taking into account the document itself (i.e., after reviewing the document), each human evaluator ranks each sentence in the set of sentences in order of most salient to least salient. In an example, the most salient sentence may receive the highest score (e.g., 15 out of 15), and the least salient sentence may receive the lowest score (e.g., 1 out of 15). Any other ranking or scoring system may be used as long as each of the sentences is assigned a value that indicates its ranking. Using the above example, each human evaluator is provided with the set of 15 sentences, and ranks the 15 sentences in order of which sentence he/she deems is the most significant (rating it a 15) to the least significant (rating it a 1).

[0022] By ranking only those sentences within the set of sentences, the human evaluators are not ranking sentences that have been deemed functionally irrelevant by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In other words, the human evaluators are not ranking sentences that were not ranked by any one of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

[0023] The human evaluation data (i.e., the human rank of each sentence in the set) is then input into the local computing system 12 using the input device 22, and is received by the processor 16. The processor 16 executes computer readable instructions 18 to assign a relative human rank to each of the sentences in the set based upon the human evaluation data (i.e., the human rankings). The relative human rank may be a weight that is assigned to each sentence evaluated by the human evaluators, where the assigned relative human rank/weight depends upon the rank of each individual sentence when all of the received human evaluation data is taken into account. For example, if the processor 16 receives human evaluation data from one human evaluator, the relative human rank/weight may match the human ranking. As an example, the sentences ranked 15 (most salient) through 1 (least salient) by a single human evaluator may be assigned relative

human ranks/weights, respectively, of 15.0 (most salient) through 1.0 (least salient).

[0024] If, however, the processor 16 receives human evaluation data from more than one human evaluator, the relative human rank/weight will be based upon all of the human rankings that are received. For example, if the processor 16 receives human evaluation data from five human evaluators, the human rankings for each sentence will be averaged and the sentences will be re-ranked based upon the averages. The data from different human evaluators could be weighted differently, for example, if one set of human evaluation data were from an expert in a particular subject matter, this data could be weighted higher than the other sets of human evaluation data. In this instance, the re-ranking will take into account the different weights. The relative human rank/weight will then be assigned to the re-ranked sentences. The following example involves a three sentence set from an email that was evaluated by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, and 24<sub>C</sub> and a human ranking scheme of 3 = most salient, 2 = less salient, and 1 = least salient. In this example, sentence number 1 is given, by the respective human evaluators, four scores of 3 and one score of 2 (average score = 2.8), sentence number 2 is given, by the respective human evaluators, two scores of 2 and three scores of 1 (average score = 1.4), and sentence number 3 is given, by the respective human evaluators, two scores of 1 and three scores of 2 (average score = 1.6). When re-ranked based upon the average scores, the order of the sentences will be sentence number 1 with the highest average score, followed by sentence 3, and then sentence 2 with the lowest average. In this example, sentence number 1 may be assigned a relative human rank of 3.0, sentence number 3 may be assigned a relative human rank of 2.0, and sentence number 2 may be assigned a relative human rank of 1.0. As such, the relative human rank/weight assigned by the processor 16 may or may not be the same as any individual human ranking that is initially received at the processor 16 by the respective human evaluators.

[0025] The relative human rank/weight may correspond with the number of sentences in any given set. For example, if the set contains X number of

sentences, the relative human rank/weight may range from X.0 down to 1.0, where X.0 is assigned to the sentence with the highest human rank or average human rank and 1.0 is assigned to the sentence with the lowest human rank or average human rank. In another example, the human evaluator may be given a set number of points (e.g., 100 points), and asked to assign the points in any manner to the sentences in the given set. Some evaluators may distribute the points relatively evenly, while other evaluators may divide all the points among a few sentences within the set. Table 4 below illustrates an example of the relative human rank/weight assigned by the processor 16 after averaging received human evaluation data for the example given herein with the set of 15 sentences that includes sentence numbers 1, 3, 4, 5, 6, 7, 9, 14, 19, 24, 25, 29, 30, 33, and 35 (originally selected and ranked by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>).

<b>Table 4</b>	
<b>Sentence Number</b>	<b>Relative Human Rank/Weight</b>
1	15.0
7	14.0
3	13.0
14	12.0
4	11.0
6	10.0
9	9.0
19	8.0
33	7.0
35	6.0
5	5.0
25	4.0
29	3.0
24	2.0
30	1.0

[0026] The processor 16, running computer readable instructions 18, is capable of determining a combined weight for each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In determining the combined weight, the processor 16 uses the weight assigned to a particular sentence from a particular summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> and uses the relative human rank assigned to the particular sentence. More

particularly, the combined weight is equal to the weight of each individual sentence selected by the particular summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> multiplied by the relative human rank assigned to that sentence. The weight of each individual sentence may also be referred to as the weight that is associated with a particular rank (e.g., ranks 1 through 10 in the output summaries in Tables 1-3) assigned by the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In an example then, the combined weight (CW) for any individual sentence (S) selected by any individual summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be determined by the following equation:

$$CW(S) = W(i) \times RHR(S(i,j)) \quad (\text{Equation 1})$$

where  $W(i)$  is the weight associated with the particular rank (i) that the sentence (S) is assigned by the summarization engine (j), and  $RHR(S(i,j))$  is the relative human rank of the sentence (S) identified by the rank and the summarization engine, or (i,j).

[0027] (j) = 1...y, where Y is the total number of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. In this example, the total number of summarization engines is 3, where engine 24<sub>A</sub> is the first engine, 24<sub>B</sub> is the second engine, and 24<sub>C</sub> is the third engine. The sentence (S) may be identified using the data in any of the output summaries. As examples, S(5, 3) is the sentence ranked 5th by summarization engine 24<sub>C</sub>, or sentence number 7 in Table 3, and S(8, 1) is the sentence ranked 8th by summarization engine 24<sub>A</sub>, or sentence number 33 in Table 1. Using Tables 1 and 4 above, the combined weight for sentence 4 selected by the summarization engine 24<sub>A</sub> is 10.0 (the weight assigned to rank 1) multiplied by 11.0 (the relative human rank assigned to sentence 4), or 110.0.

[0028] The total weight for each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be determined by adding together all of the combined weights for the sentences selected by a particular summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. As such, the total weight (TW) for any summarization engine (j) may be determined by the following equation:

$$TW_j = \sum_{i=1}^{N_s} W(i) \times RHR(S(i,j)) \quad (\text{Equation 2})$$

where  $N_s$  is the number of sentences.  $W(i)$  and  $RHR(S(i,j))$  are described above in reference to equation 1.

[0029] Table 5 below illustrates the combined weights of the individual sentences selected by each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, and also the total weight for each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. The ranks, sentence number associated with each of the ranks, and weight associated with each of the ranks correspond with the output summary for each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

<b>Rank of the Sentence (Weight)</b>	<b>Summarization Engine 24<sub>A</sub></b>		<b>Summarization Engine 24<sub>B</sub></b>		<b>Summarization Engine 24<sub>C</sub></b>	
	<u>Sentence Number</u>	<u>Combined Weight</u>	<u>Sentence Number</u>	<u>Combined Weight</u>	<u>Sentence Number</u>	<u>Combined Weight</u>
1 (10.0)	4	110.0	7	140.0	1	150.0
2 (9.0)	7	126.0	1	135.0	14	108.0
3 (8.0)	3	104.0	6	80.0	4	88.0
4 (7.0)	14	84.0	29	21.0	3	91.0
5 (6.0)	25	24.0	19	48.0	7	84.0
6 (5.0)	9	45.0	4	55.0	19	40.0
7 (4.0)	1	60.0	5	20.0	25	16.0
8 (3.0)	33	21.0	3	39.0	30	3.0
9 (2.0)	19	16.0	33	14.0	9	18.0
10 (1.0)	35	6.0	14	12.0	24	2.0
<b>Total Weight</b>		596.0		564.0		600.0

[0030] The total weights for the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be compared to one another to determine which summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> has the highest overall weight and thus is the best summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> for the document. The processor 16 running computer readable instructions 18 may perform this comparison and provide the result on a display screen of the

local computing system 12, as a printout, or in any other suitable manner. A user may also manually perform the comparison when he/she is provided with the total weights from the local computing system 12. The results in Table 5 indicate that for the document, summarization engine 24<sub>C</sub> is the best of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> with an overall weight of 600, but that summarization engine 24<sub>A</sub> is not much different with an overall weight of 596. These results may contribute to a determination as to which summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> to use in the future. For example, if summarization engine 24<sub>A</sub> were cheaper than summarization engine 24<sub>C</sub>, one may selected summarization engine 24<sub>A</sub> since its performance was similar to summarization engine 24<sub>C</sub>.

[0031] The combinatorial summarization method previously described herein provides relative, quantitative and comparative data. For larger sample sets of the summarization engines, the relative differences in total weights may indicate that some summarization engines are more or less interchangeable with one another in quality and value, and may also indicate that other summarization engine(s) should not be used in place of the summarization engine(s) having the highest total weight.

[0032] The previously described method provides a complete ranking of all of the sentences in the set of sentences selected by the various summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. Different combinations of two or more summarization engines may be evaluated using meta-algorithmic patterns, such as a voting pattern or a weighted voting pattern. Another suitable meta-algorithmic pattern that may be used includes constrained substitution, in which a suitable engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is selected for a given bandwidth, processing time, cost, or other system constraint. These patterns may be applied to the sentences within the set to further evaluate the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, or more particularly any combinations of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, that originally select, rank, and weight the sentences.

[0033] The voting pattern may be applied by the processor 16 running computer readable instructions 18. When the voting pattern is applied, the weights assigned

to a sentence by the individual summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> are added together to generate a sum of weights for that sentence. The sum of weights for any given sentence is the overall weight given by the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. A respective sum is generated for each of the sentences in the set of sentences for the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> being evaluated. For example, if it were desirable to further evaluate summarization engines 24<sub>A</sub> and 24<sub>C</sub>, the sum of weights could be calculated for the set of sentences selected by those two engines alone. In other words, the voting pattern may be applied on any number of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> ranging from two of the engines to all of the engines. It may be undesirable to exclude certain engines for any number of reasons, including cost, licensing, performance, accuracy, etc. Table 6 illustrates the sums that are generated at this step of the voting pattern for the sentences in the set selected by summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. When a sentence in the set has not been assigned a rank and an associated weight by one or more of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, the weight is N/A or zero.

<b>Sentence Number</b>	<b>Sentence Weight given by Summarization Engine 24<sub>A</sub></b>	<b>Sentence Weight given by Summarization Engine 24<sub>B</sub></b>	<b>Sentence Weight given by Summarization Engine 24<sub>C</sub></b>	<b>Sum of Weights</b>
1	4.0	9.0	10.0	23.0
3	8.0	3.0	7.0	18.0
4	10.0	5.0	8.0	23.0
5	N/A	4.0	N/A	4.0
6	N/A	8.0	N/A	8.0
7	9.0	10.0	6.0	25.0
9	5.0	N/A	2.0	7.0
14	7.0	1.0	9.0	17.0
19	2.0	6.0	5.0	13.0
24	N/A	N/A	1.0	1.0
25	6.0	N/A	4.0	10.0
29	N/A	7.0	N/A	7.0
30	N/A	N/A	3.0	3.0
33	3.0	2.0	N/A	5.0
35	1.0	N/A	N/A	1.0

[0034] At least some of the sentences are then re-ranked by the processor 16 running computer readable instructions 18 based upon the sum of weights. In an example, a predetermined number of the sentences are re-ranked, and the predetermined number forms a discrete subset of sentences. The discrete subset of sentences may include the same number of sentences that each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> originally selected. For example, if each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> originally selected 10 sentences, then the discrete subset of sentences that are re-ranked based upon the sum of weights may include 10 sentences. In another example, the processor 16 may re-rank all of the sentences in the set (which includes all of the sentences selected by the various summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>), and thus the discrete subset may include more sentences than any one summarization engine individually selected 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

[0035] When the sentences are re-ranked, the sentences are also assigned a new weight based upon the re-ranking. The new weight is representative of the weight assigned by the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. Any of the previously described ranking and weighting schemes may be used to re-rank and re-weight the sentences. For example, the sentence with the highest sum of weights may be ranked first (i.e., the most salient sentence) and assigned the highest new weight of the re-ranked sentences. When two of the sentences have the same sum, the processor 16 may be configured to review the relative human ranking to determine which of the sentences should be ranked and weighted higher among the re-ranked sentences. For example, the sentence with the highest overall ranking by the human evaluators may be selected. Any other tie-breaking scheme may be used to rank the sentences having the same sum.

[0036] After re-ranking the sentences based upon the respective sums and assigning the new weights to the re-ranked sentences, the processor 16 running computer readable instructions 18 is capable of determining a total weight of the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. The total weight of the

combination is based upon the new weights and the previously assigned relative human ranks of each sentence in the discrete subset (i.e., the re-ranked sentences). To determine the total weight of the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, the processor 16, running computer readable instructions 18, may first determine a new combined weight for each of the re-ranked sentences. In determining the new combined weight, the processor 16 multiplies the new weight assigned to a particular sentence by the relative human rank previously assigned to the particular sentence. In an example then, the new combined weight (NCW) for any individual sentence (S) that has been re-ranked according to the sum of weights may be determined by the following equation:

$$NCW(S) = NW(i) \times RHR(i) \quad (\text{Equation 3})$$

where  $NW(i)$  is the new weight associated with the particular rank (i) that the re-ranked sentence is assigned, and  $RHR(i)$  is the relative human rank previously assigned to the sentence.

[0037] The total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be determined by adding together all of the new combined weights for the re-ranked sentences. As such, the total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (i.e.,  $TW_{\text{combo}}$ ) may be determined from the following equation:

$$TW_{\text{combo}} = \sum_{i=1}^{N_S} NW(i) \times RHR(i) \quad (\text{Equation 4})$$

where  $N_S$  is the number of sentences.  $NW(i)$  and  $RHR(i)$  are described above in reference to equation 3.

[0038] Table 7 below illustrates the top 10 re-ranked sentences based upon the sum of weights results shown in Table 6. Table 7 also illustrates the new combined weights of the individual re-ranked sentences, and also the total weight for the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

<b>Table 7</b>			
<b>New Rank</b>	<b>New Weight</b>	<b>Sentence Number (sum of weights from Table 6)</b>	<b>(New Weight) x (Relative Human Ranking) = New Combined Weight</b>
1	10.0	7 (25.0)	10.0 x 14.0 = 140.0
2	9.0	1 (23.0)	9.0 x 15.0 = 135.0
3	8.0	4 (23.0)	8.0 x 11.0 = 88.0
4	7.0	3 (18.0)	7.0 x 13.0 = 91.0
5	6.0	14 (17.0)	6.0 x 12.0 = 72.0
6	5.0	19 (13.0)	5.0 x 8.0 = 40.0
7	4.0	25 (10.0)	4.0 x 4.0 = 16.0
8	3.0	6 (8.0)	3.0 x 10.0 = 30.0
9	2.0	29 (7.0)	2.0 x 3.0 = 6.0
10	1.0	9 (7.0)	1.0 x 9.0 = 9.0
<b>Total Weight for the Combination</b>			<b>627.0</b>

[0039] The maximum total weight ( $TW_{max}$ ) for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may also be determined using Equation 4. It is to be understood that the maximum total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is obtained if the re-ranked order for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> were the same as the rankings of the human evaluators. For example, when the order of the sentences in the re-ranked order corresponded exactly with the order of the sentences according to the relative human ranks/weights, the maximum total weight is obtained. In this instance, the highest new weight would be multiplied by the highest relative human rank/weight to give the highest new combined weight, the second highest new weight would be multiplied by the second highest relative human rank/weight to give the second highest new combined weight, and so on. All of these new combined weights are added together by the processor 16 running computer readable code 18 in order to calculate the maximum total weight for any combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

[0040] For the example provided herein, the maximum total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is 660.0. The total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> determined when the

voting pattern is applied and the total weight for each individual summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be compared to the maximum total weight for the combination in order to evaluate the overall performance of the individual engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> and the combination of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. For example, the total weight for the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (i.e., 627.0 shown in Table 7) is much closer to the maximum total weight of 660.0 than the total weight for any of the individual summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, which had individual total weights of 596.0, 564.0, and 600.0, respectively. In this example, the combination of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (when applying the voting pattern) is a much improved summarization engine because it is 45% closer to the maximum total weight than the best individual summarization engine 24<sub>C</sub>.

[0041] In another example, a weighted voting pattern may be applied by the processor 16 running computer readable instructions 18. When the weighted voting pattern is applied, a weighted combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is utilized to re-rank the sentences.

[0042] At the outset, a weight for each of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be calculated. In an example, the weight of each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be proportionate to the inverse of the error,  $e_j$ , of that summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, where the error is equal to the maximum total weight for the combination minus the total weight for the individual summarization engine (i.e.,  $e_j = TW_{max} - TW_j$ ). In this example, the weight ( $W_j$ ) of each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be determined by the following equation:

$$W_j = \frac{1/e_j}{\sum_{j=1}^{N_{Summ}} 1/e_j} \quad \text{(Equation 5)}$$

where  $N_{Summ}$  is the number of summarization engines. The calculation of the weight of each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> using equation 5 is shown in Table 8.

Table 8		
Summarization Engine	Error, $e_j$	Weight, $W_j$
24 <sub>A</sub>	660.0 – 596.0 = 64	$\frac{1/64}{((1/64)+(1/96)+(1/60))} = 0.366$
24 <sub>B</sub>	660.0 – 564.0 = 96	$\frac{1/96}{((1/64)+(1/96)+(1/60))} = 0.244$
24 <sub>C</sub>	660.0 – 600.0 = 60	$\frac{1/60}{((1/64)+(1/96)+(1/60))} = 0.390$

[0043] In the example disclosed herein, the weight of each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> is based upon one set of training data (i.e., data obtained for the document after one summarization and one human evaluation is performed). In other instances, however, it may be more desirable to obtain multiple sets of training data so that the error is based upon more than one set of training data. For example, multiple documents may be summarized and evaluated, and all of the data may be used to obtain an average total weight for the individual summarization engine, which in turn may be used to calculate the error.

[0044] In another example, the weight of the summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be calculated using a confidence value of that particular engine.

[0045] When the weighted voting pattern is applied, a weighted sum may be calculated for each sentence based on the respective weights of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> and the weight originally assigned to each sentence from each of the respective summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. A respective weighted sum is generated for each of the sentences in the set of sentences for the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> being evaluated. For example, if it were desirable to further evaluate summarization engines 24<sub>A</sub> and 24<sub>C</sub>, the weighted sum could be calculated for the set of sentences selected by those two engines alone. As such, the weighted voting pattern may be applied to all of the engines, or any combination of the engines. The weight sum may be calculated using the following equation:

$$Weighted\ Sum = \sum_{j=1}^{N_{Summ}} W_j \times W_i \quad (Equation\ 6)$$

where  $N_{\text{Summ}}$  is the number of summarization engines,  $W_j$  is weight of each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, and  $W_i$  is the weight associated with the particular rank (i) that the sentence (S) is assigned by the summarization engine (j). Table 9 illustrates the weighted sums that are generated at this step of the weighted voting pattern for the sentences in the set selected by summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. When a sentence in the set has not been assigned a rank and an associated weight by one or more of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>,  $W_j \times W_i$  is N/A or zero.

<b>Table 9</b>				
<b>Sentence Number</b>	<b>Summarization Engine 24<sub>A</sub> <math>W_j \times W(i)</math></b>	<b>Summarization Engine 24<sub>B</sub> <math>W_j \times W(i)</math></b>	<b>Summarization Engine 24<sub>C</sub> <math>W_j \times W(i)</math></b>	<b>Weighted Sum</b>
1	$0.366 \times 4.0 = 1.464$	$0.244 \times 9.0 = 2.196$	$0.390 \times 10.0 = 3.9$	7.560
3	$0.366 \times 8.0 = 2.928$	$0.244 \times 3.0 = 0.732$	$0.390 \times 7.0 = 2.73$	6.390
4	$0.366 \times 10.0 = 3.66$	$0.244 \times 5.0 = 1.22$	$0.390 \times 8.0 = 3.12$	8.000
5	N/A	$0.244 \times 4.0 = 0.976$	N/A	0.976
6	N/A	$0.244 \times 8.0 = 1.952$	N/A	1.952
7	$0.366 \times 9.0 = 3.294$	$0.244 \times 10.0 = 2.44$	$0.390 \times 6.0 = 2.34$	8.074
9	$0.366 \times 5.0 = 1.83$	N/A	$0.390 \times 2.0 = 0.78$	2.610
14	$0.366 \times 7.0 = 2.562$	$0.244 \times 1.0 = 0.244$	$0.390 \times 9.0 = 3.51$	6.316
19	$0.366 \times 2.0 = 0.732$	$0.244 \times 6.0 = 1.464$	$0.390 \times 5.0 = 1.95$	4.146
24	N/A	N/A	$0.390 \times 1.0 = 0.390$	0.390
25	$0.366 \times 6.0 = 2.196$	N/A	$0.390 \times 4.0 = 1.56$	3.756
29	N/A	$0.244 \times 7.0 = 1.708$	N/A	1.708
30	N/A	N/A	$0.390 \times 3.0 = 1.17$	1.17
33	$0.366 \times 3.0 = 1.098$	$0.244 \times 2.0 = 0.488$	N/A	1.586
35	$0.366 \times 1.0 = 0.366$	N/A	N/A	0.366

[0046] At least some of the sentences are then re-ranked by the processor 16 running computer readable instructions 18 based upon the weighted sums. In an example, a predetermined number of the sentences are re-ranked, and the predetermined number forms a discrete subset of sentences. The discrete subset of sentences may include the same number of sentences that each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> originally selected. For example, if each summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> originally selected 10 sentences, then the discrete subset of sentences that are re-ranked based upon the weighted sums may include 10 sentences. In another example, the processor 16 may re-rank all of the sentences in the set (which includes all of the sentences selected by the various summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>), and thus the discrete subset may include more sentences than any one summarization engine individually selected 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

[0047] When the sentences are re-ranked, the sentences are also assigned a new weight based upon the re-ranking. Any of the previously described ranking and weighting schemes may be used to re-rank and re-weight the sentences. For example, the sentence with the highest weighted sum may be ranked first (i.e., the most salient sentence) and assigned the highest new weight of the re-ranked sentences. When two of the sentences have the same weighted sum, the processor 16 may be configured to review the relative human ranking to determine which of the sentences should be ranked and weighted higher among the re-ranked sentences. For example, the sentence with the highest ranking among the relative human ranks may be selected. Other tie-breaking schemes may also be used.

[0048] After re-ranking the sentences based upon the respective weighted sums and assigning the new weights to the re-ranked sentences, the processor 16 running computer readable instructions 18 is capable of determining a total weight of the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. The total weight of the combination is based upon the new weights and the previously assigned

relative human ranks of each sentence in the discrete subset (i.e., the re-ranked sentences). To determine the total weight of the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, the processor 16, running computer readable instructions 18, may first determine a new combined weight for each of the re-ranked sentences. In determining the new combined weight, the processor 16 multiplies the new weight assigned to a particular sentence by the relative human rank previously assigned to the particular sentence. The new combined weight when the weighted voting pattern is applied may be determined using Equation 3, except that NW(i) is the new weight associated with the particular rank (i) that the sentence is assigned when re-ranked based upon the weighted sums.

[0049] The total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be determined by adding together all of the new combined weights for the sentences re-ranked based upon the weighted sums.

[0050] Table 10 below illustrates the top 10 re-ranked sentences based upon the sums shown in Table 9. Table 10 also illustrates the new combined weights of the individual re-ranked sentences, and also the total weight for the combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>.

<b>Table 10</b>			
<b>New Rank</b>	<b>New Weight</b>	<b>Sentence Number (weighted sums from Table 9)</b>	<b>(New Weight) x (Relative Human Ranking) = New Combined Weight</b>
1	10.0	7 (8.074)	10.0 x 14.0 = 140.0
2	9.0	4 (8.000)	9.0 x 11.0 = 99.0
3	8.0	1 (7.560)	8.0 x 15.0 = 120.0
4	7.0	3 (6.390)	7.0 x 13.0 = 91.0
5	6.0	14 (6.316)	6.0 x 12.0 = 72.0
6	5.0	19 (4.146)	5.0 x 8.0 = 40.0
7	4.0	25 (3.756)	4.0 x 4.0 = 16.0
8	3.0	9 (2.610)	3.0 x 9.0 = 27.0
9	2.0	6 (1.952)	2.0 x 10.0 = 20.0
10	1.0	29 (1.708)	1.0 x 3.0 = 3.0
<b>Total Weight</b>			<b>628.0</b>

[0051] The total weight for the combination of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> determined when the weighted voting pattern is applied and the total weight for each individual summarization engine 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> may be compared to the maximum total weight for the combination in order to evaluate the overall performance of the individual engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> and the combination of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>. For example, the total weight for the weighted combination of the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (i.e., 628.0 shown in Table 10) is much closer to the maximum total weight of 660.0 than the total weight for any of the individual summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub>, which had individual total weights of 596.0, 564.0, and 600.0, respectively. In this example, the combination of the engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> (when applying the weighted voting pattern) is a much improved summarization engine because it is 46.7% closer to the maximum total weight than the best individual summarization engine 24<sub>C</sub>.

[0052] Referring now to Fig. 2, the combinatorial summarizer 10' is implemented in the cloud computing system 14. The cloud computing system or "cloud" is a computer network accessible over the Internet that is dynamically scalable with virtualized resources, distributed systems, shared resources, local area networks (LAN), or an Intranet in the case of a private cloud. In an example, the cloud 14 may or may not be physically distributed. The user of the combinatorial summarizer 10' not required to have knowledge or expertise in the infrastructure of the cloud 14 that relies on the Internet to satisfy his/her summarization and evaluation needs. The cloud 14 provides the services of the combinatorial summarizer 10' that are accessible from a web browser, while computer readable instructions 18 and data may be stored on servers (which include media 20 and memory) in the cloud 14.

[0053] The cloud computing system 14 may be a private cloud, a public cloud or a hybrid cloud. Further, the cloud 14 may be a combination cloud computing system including a private cloud (or multiple private clouds) and a public cloud (or multiple public clouds).

[0054] The cloud computing system 14 may include multiple pieces of hardware operatively coupled over the network so that they can perform a specific computing task (e.g., the combinatorial summarization methods disclosed herein). The cloud 14 may include a combination of physical hardware 30, software 32, and virtual hardware 34.

[0055] The physical hardware 30 may include, among others, processors, memory devices, and networking equipment. The virtual hardware 34 is a type of software that is processed by the physical hardware 30 and designed to emulate specific hardware. As an example, virtual hardware 34 may include a virtual machine (VM), i.e., a software implementation of a computer that supports execution of an application like a physical machine.

[0056] An application, as used herein, refers to a set of specific instructions executable by a computing system for facilitating carrying out a specific task. For example, an application may take the form of a web-based tool providing users with a specific functionality, e.g., document summarization and/or evaluation. Software 32 is a set of instructions and data configured to cause virtual hardware 34 to execute the application. The cloud computing system 14 can thus render a particular application available to users associated with the computing system 26.

[0057] Executing the application in the cloud 14 may involve receiving a number of requests, processing the requests according to the particular functionality implemented by the application, and returning request responses to the requesting computing system 26. For executing the application, the resources (e.g., physical hardware 30, virtual hardware 34, and software 32) of the cloud computing system 14 may be scaled depending on the demands posed on the application. For example, cloud 14 may vary the size of the resources allocated to the application depending on the number of requests, the number of users interacting with the application, or requirement on the performance of the application (e.g., a maximum response time). While not shown, it is to be understood that the cloud 14 may also include an interface that allows the computing system 26 to communicate with the components of the cloud 14.

[0058] Referring still to Fig. 2, the physical hardware 30 of the cloud 4 may include the processor 16 and the medium 20. The processor 16 may be any processor that is capable of executing program instructions 18 stored in the media 20 to implement, for example, the combinatorial summarization methods as disclosed herein. The media 20 may include an operating system and applications, such as a combinatorial summarization application. The operating system may be a collection of programs that, when executed by the processor 16, serves as a platform on which the combinatorial summarization application can run. Some examples of operating systems include various versions of Linux® and Microsoft Windows®.

[0059] As illustrated in Fig. 2, the computing system 26 may be connected to the cloud computing system 14 via a link 28. It is to be understood that the computing system 26 may include one or more standalone computing systems and/or a network of computing systems. The links 28 may be one or more of cable, wireless, fiber optic, or remote connections via a telecommunication link, an infrared link, a radio frequency link, or any other connectors or systems that provide electronic communication. Links 28 may include, at least in part, an intranet, the Internet, or a combination of both. The links 28 may also include intermediate proxies, routers, switches, load balancers, and the like.

[0060] The methods disclosed herein advantageously utilize human feedback in addition to weighting schemes to quantitatively evaluate the summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> alone or in combination with one another. Utilizing data from a pool of summarization engines 24<sub>A</sub>, 24<sub>B</sub>, 24<sub>C</sub> results in improved summarization.

[0061] While several examples have been described in detail, it will be apparent to those skilled in the art that the disclosed examples may be modified. Therefore, the foregoing description is to be considered non-limiting.

What is claimed is:

1. A combinatorial summarizer, comprising:  
a plurality of summarization engines, each of which is to:  
select a respective plurality of sentences; and  
5 generate a relative rank and an associated weight for each sentence  
of the respective plurality of sentences;  
a processor in selective communication with each of the plurality of  
summarization engines; and  
computer readable instructions executable by the processor and embodied  
10 on a tangible, non-transitory computer readable medium, the computer readable  
instructions including:  
computer readable instructions to determine a combined weight for  
each sentence of each of the respective plurality of sentences, the  
combined weight being based upon the respective associated weight and a  
15 respective relative human rank for each sentence in a set of sentences  
including all of the sentences of each of the respective plurality of  
sentences; and  
computer readable instructions to determine a total weight for each of  
the plurality of summarization engines based, respectively, upon the  
20 combined weights for each sentence of each of the respective plurality of  
sentences.
2. The combinatorial summarizer as defined in claim 1 wherein the  
computer readable instructions further include computer readable instructions to  
25 apply a meta-algorithmic pattern to each sentence of each of the respective  
plurality of sentences.
3. The combinatorial summarizer as defined in claim 2 wherein the meta-  
algorithmic pattern is a voting pattern, and wherein the computer readable  
30 instructions to apply the voting pattern include:

computer readable instructions to determine a sum of the associated weights for each sentence of each of the respective plurality of sentences;

computer readable instructions to rank a predetermined number of a discrete subset of sentences of each of the respective plurality of sentences

5 according to the sum of the associated weights;

computer readable instructions to assign a new weight to the sentences of the discrete subset; and

computer readable instructions to determine a total weight for a combination of the plurality of summarization engines based upon the new weight and the  
10 relative human rank for the sentences of the discrete subset.

4. The combinatorial summarizer as defined in claim 2 wherein the meta-algorithmic pattern is a weighted voting pattern, and wherein the computer readable instructions to apply the weighted voting pattern include:

15 computer readable instructions to determine a weight for each of the plurality of summarization engines;

computer readable instructions to determine a weighted sum for each sentence of each of the respective plurality of sentences, the weighted sum being based on the weights for each of the plurality of summarization engines and the  
20 associated weights for each sentence of each of the respective plurality of sentences;

computer readable instructions to rank a predetermined number of a discrete subset of sentences of each of the respective plurality of sentences according to the weighted sum;

25 computer readable instructions to assign a new weight to the sentences of the discrete subset; and

computer readable instructions to determine a total weight for a combination of the plurality of summarization engines based upon the new weight and the relative human rank for the sentences of the discrete subset.

30

5. The combinatorial summarizer as defined in claim 1, further comprising:  
an input device in selective communication with the processor, the input  
device to receive a human ranking for each sentence in the set of sentences; and  
computer readable instructions for assigning the relative human rank to  
5 each sentence in the set of sentences based upon the human ranking.

6. The combinatorial summarizer as defined in claim 1 wherein the  
summarizer is part of a cloud computing system.

10 7. The combinatorial summarizer as defined in claim 1 wherein the  
computer readable instructions further include computer readable instructions to  
generate the set of sentences.

8. A combinatorial summarization method, comprising:  
15 receiving, from each of a plurality of summarization engines and by a  
processor running computer readable instructions embodied on a tangible, non-  
transitory computer readable medium, a relative rank and an associated weight for  
each sentence of a respective plurality of sentences selected by each of the  
plurality of summarization engines;  
20 receiving, by the processor, a relative human rank for each sentence in a set  
of sentences including all of the sentences of each of the respective plurality of  
sentences;  
determining, by the processor, a combined weight for each sentence of each  
of the respective plurality of sentences, the combined weight being based upon the  
25 respective associated weight and the respective relative human rank; and  
determining, by the processor, a total weight for each of the plurality of  
summarization engines based, respectively, upon the combined weights for each  
sentence of each of the respective plurality of sentences.

9. The combinatorial summarization method as defined in claim 8, further comprising evaluating the plurality of summarization engines based upon the total weights.

5           10. The combinatorial summarization method as defined in claim 8, further comprising applying, by the processor, a meta-algorithmic pattern to each sentence of each of the respective plurality of sentences.

10           11. The combinatorial summarization method as defined in claim 10, wherein the meta-algorithmic pattern is a voting pattern, and wherein applying the voting pattern includes:

          determining, by the processor, a sum of the associated weights for each sentence of each of the respective plurality of sentences;

15           ranking, by the processor, a predetermined number of a discrete subset of sentences of each of the respective plurality of sentences according to the sum of the associated weights;

          assigning, by the processor, a new weight to the sentences of the discrete subset; and

20           determining, by the processor, a total weight for a combination of the plurality of summarization engines based upon the new weight and the relative human rank for the sentences of the discrete subset.

25           12. The combinatorial summarization method as defined in claim 10, wherein the meta-algorithmic pattern is a weighted voting pattern, and wherein applying the weighted voting pattern includes:

          determining, by the processor, a weight for each of the plurality of summarization engines;

          determining, by the processor, a weighted sum for each sentence of each of the respective plurality of sentences, the weighted sum being based on the weights

for each of the plurality of summarization engines and the associated weights for each sentence of each of the respective plurality of sentences;

ranking, by the processor, a predetermined number of a discrete subset of sentences of each of the respective plurality of sentences according to the

5 weighted sum;

assigning, by the processor, a new weight to the sentences of the discrete subset; and

determining, by the processor, a total weight for a combination of the plurality of summarization engines based upon the new weight and the relative

10 human rank for the sentences of the discrete subset.

13. A combinatorial summarization method, comprising:

determining, by a processor running computer readable instructions embodied on a tangible, non-transitory computer readable medium, i) a sum of

15 weights associated with each sentence of a respective plurality of sentences selected by a plurality of summarization engines, or ii) a weighted sum for each

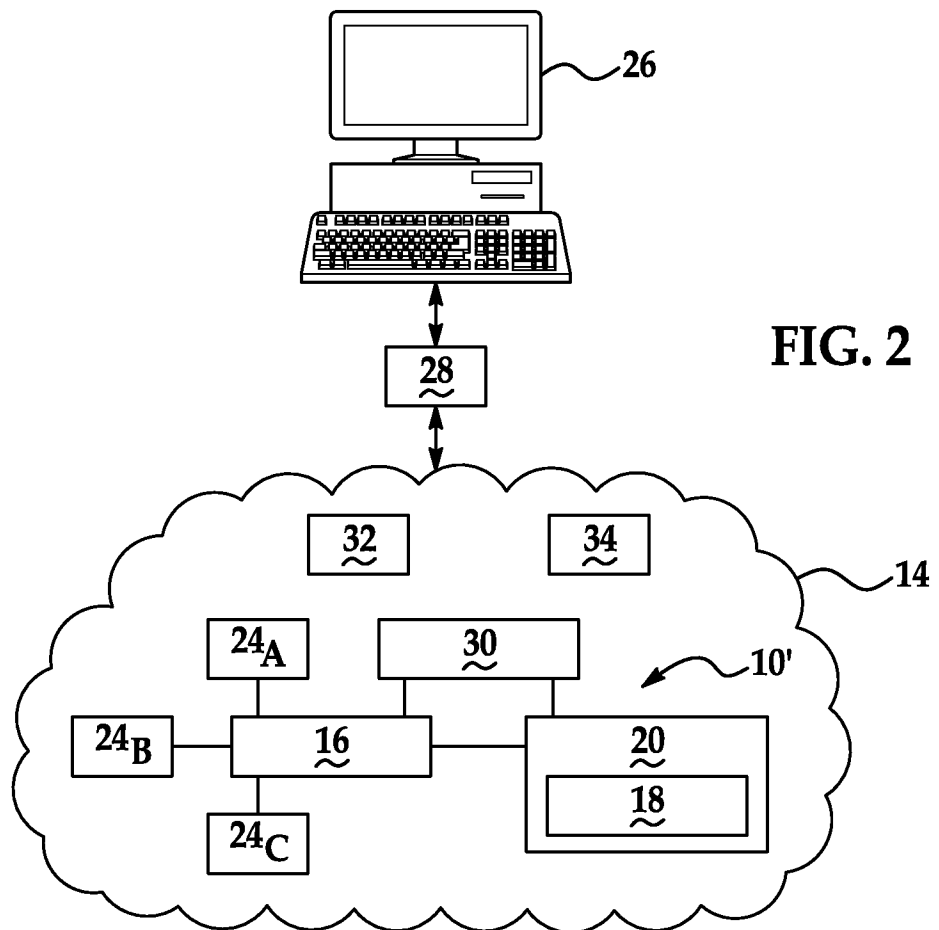
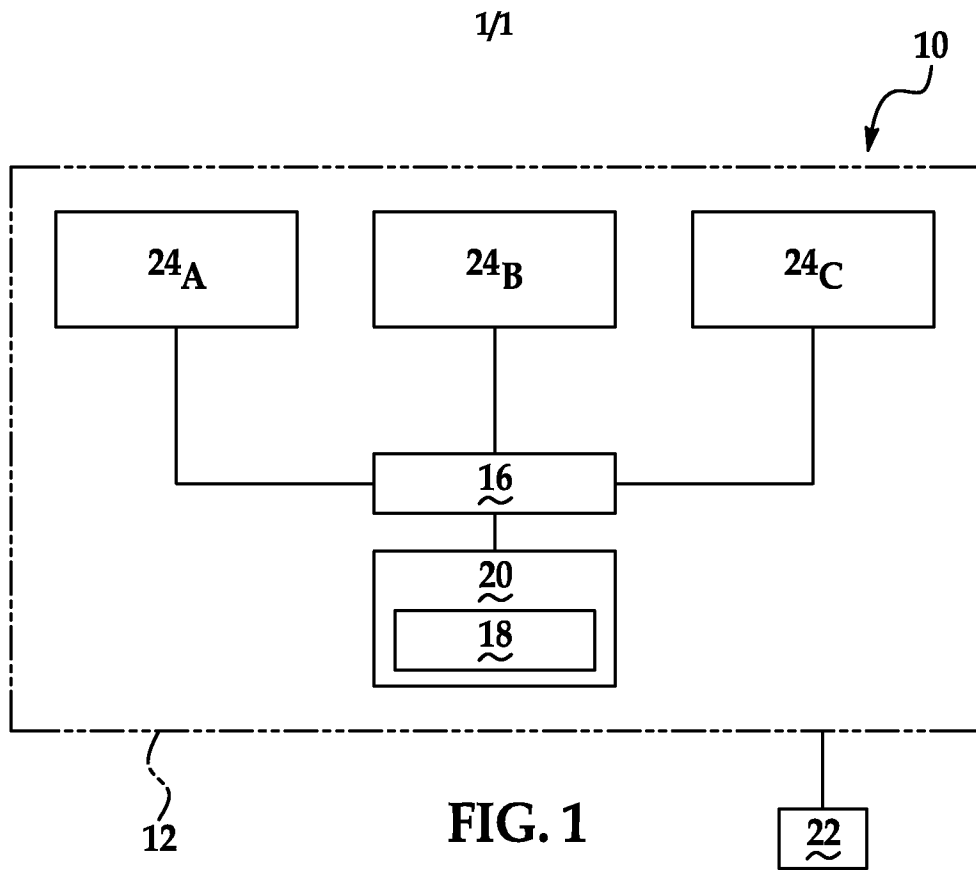
sentence of a respective plurality of sentences selected by a plurality of summarization engines, the sum being based on weights for each of the plurality of summarization engines and weights associated with each sentence of each of the

20 respective plurality of sentences;

ranking, by the processor, a predetermined number of a discrete subset of sentences of each of the respective plurality of sentences according to i) the sum of weights, or ii) the weighted sum;



25 assigning, by the processor, a new weight to the sentences of the discrete subset; and

determining, by the processor, a total weight for a combination of the plurality of summarization engines based upon the new weight and a relative human rank for the sentences of the discrete subset.



## INTERNATIONAL SEARCH REPORT

International application No.  
**PCT/US2012/059917**

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
<b>G06F 17/20(2006.01)i</b>		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) G06F 17/20; G06F 17/00; G06F 17/30; G06F 17/21; G06F 17/27		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models Japanese utility models and applications for utility models		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS(KIPO internal), Google & Keywords: summarizer, combinate, sentence, weight, rank, human rank		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7,292,972 B2 (LIN, XIAOFAN et al.) 06 November 2007 See column 1, lines 14-22; column 2, lines 56-59; column 3, lines 34-37; column 5, lines 23-54; column 6, lines 4-14; claims 1 and 16; and figures 1-2.	1-13
A	US 2012-0035912 A1 (LITVAK, MARINA et al.) 09 February 2012 See paragraphs [0002]-[0006], [0013], and [0056]-[0057]; and figures 1-2.	1-13
A	DEMIROS, IASON et al. "Sentence-based Text Summarization: Modeling and Evaluation." Proceedings of the 2nd Hellenic Conference on Artificial Intelligence. 11-12 April 2002. pp. 103-114. See pages 111-112.	1-13
A	US 8,171,026 B2 (KAWATANI, TAKAHIKO) 01 May 2012 See column 1, lines 14-27 and 43-44; column 2, lines 40-49; column 3, lines 38-54; claim 1; and figure 1.	1-13
A	US 7,886,235 B2 (BORNSTEIN, JEREMY J. et al.) 08 February 2011 See column 1, lines 20-23; column 2, lines 18-22; column 3, lines 10-35; and figures 2-5.	1-13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 29 April 2013 (29.04.2013)		Date of mailing of the international search report <b>30 April 2013 (30.04.2013)</b>
Name and mailing address of the ISA/KR  Korean Intellectual Property Office 189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea Facsimile No. 82-42-472-7140		Authorized officer NHO, Ji Myong Telephone No. 82-42-481-8528 

**INTERNATIONAL SEARCH REPORT**

International application No.

**PCT/US2012/059917**

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012-0240032 A1 (MCKEOWN, KATHLEEN R. et al.) 20 September 2012 See paragraphs [0001], [0008]-[0010], [0026], and [0033]; claim 1; and figure 6.	1-13

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/US2012/059917**

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 7,292,972 B2	06.11.2007	US 2004-153309 A1	05.08.2004
US 2012-0035912 A1	09.02.2012	None	
US 8,171,026 B2	01.05.2012	CN 1227614 C0 CN 1470047 A EP 1336174 A2 EP 1944756 A2 EP 1944756 A3 EP 1944757 A2 EP 1944757 A3 JP 2002-169834 A US 2004-068396 A1 US 2009-0216759 A1 US 7562066 B2 WO 02-41557 A2 WO 02-41557 A3	16.11.2005 21.01.2004 20.08.2003 16.07.2008 23.07.2008 16.07.2008 23.07.2008 14.06.2002 08.04.2004 27.08.2009 14.07.2009 23.05.2002 30.01.2003
US 7,886,235 B2	08.02.2011	US 05867164 A US 2002-0080196 A1 US 2006-0059442 A1 US 2006-0265666 A1 US 6424362 B1 US 7882450 B2	02.02.1999 27.06.2002 16.03.2006 23.11.2006 23.07.2002 01.02.2011
US 2012-0240032 A1	20.09.2012	None	