

CLAIMS

1. A method of enhancing an output of relevant content to facilitate analysis of the relevant content, comprising:
 - receiving at least one data feed of content to be analyzed;
 - filtering out non-relevant content, the non-relevant content including content unrelated to a subject associated with the analysis, the filtering including retaining relevant content;
 - extracting entity data corresponding to entities mentioned in the relevant content, wherein the extracted entity data corresponding to the entities includes metadata associated with the entities;
 - activity-mining the relevant content based on the extracted entity data and the metadata associated with the entities to associate the entities with activity tags;
 - identifying entity profiles in a database that match the entities, the identifying based on metadata associated with the entities and metadata of the entity profiles in the database; and
 - annotating the relevant content with at least one graphical indicator and at least one graphical user interface (GUI) control based at least in part on one or more of the entities, the matched entity profiles, and the associated activity tags.
2. The method of claim 1, wherein the at least one data feed includes at least one Rich Site Summary (RSS) feed, and wherein the content to be analyzed includes one or more of a news article, a blog, a social media post, a long form article, and an advertisement.
3. The method of claim 1 further comprising optimizing the content to be analyzed, wherein the optimizing includes parsing the content to extract portions of text from the content and removing extraneous information, the extraneous information including one of advertisements and external links.
4. The method of claim 1 further comprising de-duplicating the content, the de-duplicating including identifying content elements previously analyzed, or content elements already referenced in the database, and discarding the identified content elements.
5. The method of claim 1, wherein the filtering out the non-relevant content includes applying one or more of: a keyword filter and a natural expression pattern matching filter to the content to be analyzed.

6. The method of claim 1, wherein the filtering out the non-relevant content further includes clustering content elements based on a similarity of the content elements to each other.

7. The method of claim 6, wherein the clustering includes applying a Term Frequency-Inverse Document Frequency (TF-IDF) algorithm that includes determining a frequency of a term within a content element, and determining a frequency of the term across all content elements of the content to be analyzed.

8. The method of claim 1, wherein the extracting the entity data from the relevant content includes identifying entities mentioned in a content element, and obtaining the metadata associated with the mentioned entities.

9. The method of claim 8, wherein the identifying entities mentioned in the content element includes applying a natural expression language model to the relevant content.

10. The method of claim 8, wherein the metadata associated with a particular entity includes at least one of: an identification of the content element in which the particular entity is mentioned, a count of mentions of the particular entity in the content element, a mentioned form of the particular entity, a type of the particular entity, a normalized form of the particular entity, a location of the mention of the particular entity within the content element, a complete sentence in which the mention of the particular entity appears, and a label for a part of speech in which the mention of the particular entity appears within the content element.

11. The method of claim 1, wherein the identifying entity profiles in a database that match the extracted entities includes:

- identifying candidate profiles in the database based on a coarse similarity estimate of the candidate profiles to the extracted entities;

- calculating a refined candidate score for each candidate profile, the refined candidate score based on the metadata of the extracted entities and the metadata of each candidate profile;

- comparing the refined candidate score of each candidate profile with a threshold; and

- designating candidate profiles as a match to a particular extracted entity when the refined candidate score of the candidate profiles exceeds the threshold.

12. The method of claim 1, wherein the annotating the relevant content includes overlaying the at least one graphical indicator over each extracted entity, wherein the at least one graphical indicator over each extracted entity is of a different color.

13. The method of claim 1, wherein the at least one graphical indicator includes an indicator indicating that a particular extracted entity is unmatched with a profile entity in the database, and wherein the at least one GUI control includes a GUI control for executing an update to the database to store the particular extracted entity in the database.

14. A system for enhancing an output of relevant articles to facilitate analysis of the relevant articles, comprising:

- at least one data feed for receiving one or more articles to be analyzed;

- a server configured to:

- filter out non-relevant articles from the one or more articles, the non-relevant articles including articles unrelated to a subject associated with the analysis, wherein relevant articles are retained;

- extract entity data corresponding to entities mentioned in the relevant articles, wherein the extracted entity data corresponding to the entities includes metadata associated with the entities;

- mine the relevant articles for activities based on the extracted entity data and the metadata associated with the entities to associate the entities with activity tags;

- identify entity profiles in a database that match the entities, wherein the matching entity profiles are identified based on metadata associated with the entities and metadata of the entity profiles in the database; and

- annotate the relevant articles with at least one graphical indicator and at least one graphical user interface (GUI) control based at least in part on one or more of the entities, the matched entity profiles, and the associated activity tags;

- at least one user terminal with a graphical user interface configured to display the annotated relevant articles; and

- the database configured to store the extracted entity data for the entities and the correspondingly associated activity tags.

15. The system of claim 14, wherein the at least one data feed includes at least one Rich Site Summary (RSS) feed, and wherein the articles to be analyzed includes one or more of a news article, a blog, a social media post, a long form article, and an advertisement.

16. The system of claim 14, wherein the server is further configured to optimize the articles to be analyzed, wherein the configuration of the server to optimize includes configuration of the server to parse the articles to extract portions of text from the articles and remove extraneous information, the extraneous information including one of advertisements and external links.

17. The system of claim 14, wherein the configuration of the server to filter includes configuration of the server to:

apply one or more of: a keyword filter and a natural expression pattern matching filter to the articles to be analyzed, and

cluster articles based on a similarity of the articles to each other.

18. The system of claim 14, wherein the configuration of the server to identify entity profiles in a database that match the extracted entities includes configuration of the server to:

identify candidate profiles in the database based on a coarse similarity estimate of the candidate profiles to the extracted entities;

calculate a refined candidate score for each candidate profile, the refined candidate score based on the metadata of the extracted entities and the metadata of each candidate profile;

compare the refined candidate score of each candidate profile with a threshold; and

designate candidate profiles as a match to a particular extracted entity when the refined candidate score of the candidate profiles exceeds the threshold.

19. The system of claim 14, wherein the configuration of the server to annotate the relevant article includes configuration of the server to overlay the at least one graphical indicator over each extracted entity, wherein the at least one graphical indicator includes an indicator indicating that a particular extracted entity is unmatched with a profile entity in the database, and wherein the at least one GUI control includes a GUI control for executing an update to the database to store the particular extracted entity in the database.

20. A computer-based tool for facilitating analysis of relevant articles, the computer-based tool including non-transitory computer readable media having stored thereon computer code which, when executed by a processor, causes a computing device to perform operations comprising:

receiving an enhanced output from a server of a research system, the enhanced output including relevant articles overlaid with one or more graphical indicators and at least one graphical user interface (GUI) control, wherein the one or more graphical indicators are overlaid over at least one of: entities mentioned in the article and activities associated with the entities, wherein the enhanced output is generated by the server based on:

receiving at least one data feed of articles to be analyzed;

filtering out non-relevant articles to identify the relevant articles;

extracting entity data corresponding to the entities from the relevant articles;

mining the relevant articles based on the extracted entity data and metadata associated with the entities to associate the entities with the activity tags;

identifying entity profiles in a database that match the entities, the identifying based on the metadata associated with the entities and metadata of the entity profiles in the database; and

annotating the relevant articles with the one or more graphical indicator and the at least one GUI control based at least in part on one or more of the entities, the matched entity profiles, and the associated activity tags; and

displaying the enhanced output on a GUI.