

CLAIMS

What is claimed:

1. A method for determining technical emergence of a term, comprising:
accessing, by a system having at least one processor, a data set comprising a plurality of records having textual information, the plurality of records being from a predefined time interval having a base period and an active period;
determining, by the system, from the data set, a set of terms meeting predetermined criteria;
determining, by the system, at least one relative component value for each term for at least one of a portion of the predefined time interval;
determining, by the system, a numeric score for each term based at least on the relative component value; and
classifying, by the system, each term meeting a minimum numeric score as an emergent term;
wherein the method is configured to provide analysis and predictive information related to the emergent term.
2. The method of claim 1, further comprising determining, by the system, a plurality of relative component values for each term for one or more portions of the predefined time interval.
3. The method of claim 2, wherein the plurality of relative component values comprises: a first relative component value for each term based on a first portion of the active period of the predefined time interval; a second relative component value for each term based on a second portion of the active period of the predefined time interval; and a third relative component value for each term based on a third portion of the active period of the predefined time interval.
4. The method of claim 3, wherein determining a set of terms further comprises extracting, by the system, all n-grams up to a predetermined length.
5. The method of claim 4, wherein determining a set of terms further comprises reducing, by the system, a size of a term set by applying at least one filter.
6. The method of claim 4, wherein the at the least one filter includes at least one of:

- eliminating, by the system, candidate terms that are present in a controlled vocabulary;
 - applying, by the system, a general fuzzy matching algorithm;
 - applying, by the system, a folding routine; and
 - eliminating, by the system, candidate terms comprising a stopword.
7. The method of claim 1, wherein the predetermined criteria comprises at least one of:
 - at least one term instance in a minimum number of records;
 - at least one term instance in at least one record for a minimum number of record periods;
 - at least one term instance in a maximum percentage of records from a predefined first portion of the time interval of the data set;
 - a ratio of a number of records with at least one term instance from a predefined first portion of the time interval of the data set to a number of records with at least one term instance of the same term from a predefined second portion of the time interval of the data set of at least about 1:2; and
 - at least one term instance in a plurality of records having different author data.
 8. The method of claim 1, wherein determining at least one relative component value comprises assigning, by the system, a numeric value for each relative component value based on at least one trend analysis for the term.
 9. The method of claim 8, wherein the trend analysis comprises trend analysis of time series data.
 10. The method of claim 1, wherein determining at least one relative component value comprises assigning, by the system, a numeric value for each relative component value based on at least one of the following:
 - publication activity of a last portion of an active period minus publication activity of a first portion of the active period;
 - publication activity of a first recent portion of the active period minus publication activity of an immediate prior portion of the active period;
 - slope from a midpoint portion of the active period to a second recent portion of the active period; and
 - slope from a first point portion of the active period to the mid-point portion of the active period.

- 11.** The method of claim 1, wherein determining at least one relative component value comprises assigning, by the system, a numeric value for each relative component value based on at least one of the following:
- a) the difference of:
 - i) (number of records having at least one term instance from a last portion of an active period divided by the square root of the total number of records from said last portion of the active period); and
 - ii) (number of records having at least one term instance from a first portion of the active period divided by the square root of the total number of records from said first portion of the active period);
 - b) the difference of:
 - i) (number of records having at least one term instance from a first recent portion of an active period divided by the square root of the total number of records from said first recent portion of the active period); and
 - and
 - ii) (number of records having at least one term instance from an immediate prior portion of the active period divided by the square root of the total number of records from said immediate prior portion of the active period); and
 - c) the quotient of:
 - i) the difference of:
 - 1. (number of records having at least one term instance from a midpoint portion of the active period divided by the square root of the total number of records from said midpoint portion); and
 - 2. (number of records having at least one term instance from a second recent portion of the active period divided by the square root of the total number of records from said second recent portion); and
 - ii) change in time from said midpoint portion and said second recent portion.
- 12.** The method of claim 1, wherein determining at least one relative component value comprises assigning, by the system, a numeric value for each relative component value based on at least one of the following:

- change from the most recent 3-year period of an active period to the first 3-year period of the active period;
 - change from the most recent 2-year period of the active period to the prior 2-year period;
 - slope from mid-year of the active period to the most recent year of the active period; and
 - slope from first year of the active period to mid-year of the active period.
- 13.** The method of claim 1, wherein determining a numeric score further comprises applying, by the system, a predetermined weight to each relative component value to determine the numeric score.
- 14.** The method of claim 1, wherein determining a numeric score further comprises at least one of: applying, by the system, a predetermined weight to each relative component value; and aggregating the relative component values.
- 15.** The method of claim 14, further comprising generating, by the system, at least one indicator for a set of records sharing at least one feature, the at least one indicator being based on at least one of:
- summation of scores of terms having scores meeting a predetermined threshold;
 - number of instances of terms having scores meeting a predetermined threshold;
 - average score of terms having scores meeting a predetermined threshold for set of records;
 - summation of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - average of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - summation of a square root of scores of terms having scores meeting a predetermined threshold;
 - average of a square root of scores of terms having scores meeting a predetermined threshold;
 - number of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;

- percentage of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;
average score of terms having scores meeting a predetermined threshold and appearing within a record;
percentage of records within the set meeting a predetermined average score;
percentage of records within the set having a predetermined number of instances of terms having scores meeting a predetermined threshold and appearing within a record;
number of unique terms having scores meeting a predetermined threshold; and
number of records meeting a predetermined score summation threshold.
16. The method of claim 15, wherein generating at least one indicator further comprises assigning a value to each indicator.
 17. The method of claim 16, wherein assigning a value comprises calculating a numeric summation of the square root of scores of terms having scores meeting a predetermined threshold for each first instance of the term within each record of the set.
 18. The method of claim 17, further comprising normalizing the numeric summation based on: the number of records in the set or square root of the number of records in the set.
 19. The method of claim 15, wherein the feature relates to information associated with at least one of the following:
 - a person; an organization; a country; a data source; a document type; a document section; a technology; a technical domain; a technical field; a time interval; a geographical location; and a patent class.
 20. The method of claim 19, further comprising displaying, by the system, the emergent terms comprising the numeric scores.
 21. The method of claim 20, further comprising displaying, by the system, the at least one indicator.
 22. The method of claim 20, further comprising displaying, by the system, a plurality of indicators for a plurality of record sets.
 23. The method of claim 1, wherein the data set is received from a database.
 24. The method of claim 1, wherein the records comprise documents.

25. The method of claim 24, wherein the documents comprise research publications or patent publications, or a combination thereof.
26. The method of claim 1, wherein the terms comprise abstract phrases or title phrases, or a combination thereof.
27. The method of claim 1, wherein the base period and the active period comprise the total time interval.
28. The method of claim 7, wherein the base period of the time interval comprises from about 1% to about 50% of the total time interval.
29. The method of claim 7, wherein the active period of the time interval comprises from about 50% to about 99% of the total time interval.
30. The method of claim 7, wherein the first portion of the time interval comprises from greater than 0 days to about 100 years.
31. The method of claim 7, wherein the second portion of the time interval comprises from greater than 0 days to about 100 years
32. The method of claim 7, wherein the first portion of the time interval comprises from about 1% to about 50% of the total time interval.
33. The method of claim 7, wherein the second portion of the time interval comprises from about 50% to about 99% of the total time interval.
34. The method of claim 7, wherein the first portion of the time interval comprises from about a first 1% to about a first 50% of the total time interval.
35. The method of claim 7, wherein the second portion of the time interval comprises from about a second 50% to about a second 99% of the total time interval.
36. A computer-readable non-transitory storage medium having a set of computer-accessible instructions which when executed, cause at least one processor to perform operations comprising:
 - accessing a data set comprising a plurality of records having textual information, the plurality of records being from a predefined time interval having a base period and an active period;
 - determining from the data set, a set of terms meeting predetermined criteria;
 - determining at least one relative component value for each term for at least one of a portion of the predefined time interval;
 - determining a numeric score for each term based on the relative component value; and

- classifying each term meeting a minimum numeric score as an emergent term.
37. The medium of claim 36, wherein the operations further comprise determining a plurality of relative component values for each term for one or more portions of the predefined time interval.
38. The medium of claim 36, wherein the plurality of relative component values comprises: a first relative component value for each term based on a first portion of the active period of the predefined time interval; a second relative component value for each term based on a second portion of the active period of the predefined time interval; and a third relative component value for each term based on a third portion of the active period of the predefined time interval.
39. The medium of claim 36, wherein determining a set of terms further comprises reducing a size of a term set by applying at least one filter.
40. The medium of claim 39, wherein the at the least one filter includes at least one of: eliminating candidate terms that are present in a controlled vocabulary; applying a general fuzzy matching algorithm; applying a folding routine; and eliminating candidate terms comprising a stopword.
41. The medium of claim 36, wherein the predetermined criteria comprises at least one of: at least one term instance in a minimum number of records; at least one term instance in at least one record for a minimum number of record periods; at least one term instance in a maximum percentage of records from a predefined first portion of the time interval of the data set; a ratio of a number of records with at least one term instance from a predefined first portion of the time interval of the data set to a number of records with at least one term instance of the same term from a predefined second portion of the time interval of the data set of at least about 1:2; and at least one term instance in a plurality of records having different author data.
42. The medium of claim 36, wherein determining at least one relative component value comprises assigning a numeric value for each relative component value based on at least one trend analysis for the term.
43. The medium of claim 42, wherein the trend analysis comprises analysis of time series data.

44. The medium of claim 36, wherein determining at least one relative component value comprises assigning a numeric value for each relative component value based on at least one of the following: publication activity of a last portion of an active period minus publication activity of a first portion of the active period; publication activity of a first recent portion of the active period minus publication activity of an immediate prior portion of the active period; slope from a midpoint portion of the active period to a second recent portion of the active period; and slope from a first point portion of the active period to the mid-point portion of the active period.
45. The medium of claim 36, wherein determining a numeric score further comprises applying a predetermined weight to each relative component value to determine the numeric score.
46. The medium of claim 36, wherein determining a numeric score further comprises at least one of: applying a predetermined weight to each relative component value; and aggregating the relative component values.
47. The medium of claim 36, wherein the operations further comprise generating at least one indicator for a set of records sharing at least one feature, the at least one indicator being based on at least one of:
 - summation of scores of terms having scores meeting a predetermined threshold;
 - number of instances of terms having scores meeting a predetermined threshold;
 - average score of terms having scores meeting a predetermined threshold for set of records;
 - summation of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - average of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - summation of a square root of scores of terms having scores meeting a predetermined threshold;
 - average of a square root of scores of terms having scores meeting a predetermined threshold;

- number of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;
 - percentage of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;
 - average score of terms having scores meeting a predetermined threshold and appearing within a record;
 - percentage of records within the set meeting a predetermined average score;
 - percentage of records within the set having a predetermined number of instances of terms having scores meeting a predetermined threshold and appearing within a record; and
 - number of unique terms having scores meeting a predetermined threshold;
 - number of records meeting a predetermined score summation threshold.
48. The medium of claim 47, wherein the operations further comprise generating at least one indicator further comprises assigning a value to each indicator.
49. The medium of claim 48, wherein assigning a value comprises calculating a numeric summation of the square root of scores of terms having scores meeting a predetermined threshold for each first instance of the term within each record of the set.
50. The medium of claim 49, wherein the operations further comprise normalizing the numeric summation based on the number of records in the set, or square root of the number of records in the set.
51. The medium of claim 47, wherein the feature relates to information associated with at least one of the following:
- a person; an organization; a country; a data source; a document type; a document section; a technology; a technical domain; a technical field; a time interval; a geographical location; and a patent class.
52. The medium of claim 51, wherein the operations further comprise displaying the emergent terms comprising the numeric scores.
53. The medium of claim 51, wherein the operations further comprise displaying the at least one indicator.
54. The medium of claim 53, wherein the operations further comprise displaying a plurality of indicators for a plurality of record sets.
55. A system, comprising:

- at least one memory having computer-accessible instructions; and
at least one processor functionally coupled to the at least one memory and configured by at least a portion of the computer-accessible instructions to:
access a data set comprising a plurality of records having textual information, the plurality of records being from a predefined time interval having a base period and an active period;
determine from the data set, a set of terms meeting predetermined criteria;
determine at least one relative component value for each term for at least one of a portion of the predefined time interval;
determine a numeric score for each term based on the relative component value; and
classify each term meeting a minimum numeric score as an emergent term.
- 56.** The system of claim 55, wherein the at least one relative component value comprises a plurality of relative component values for each term for one or more portions of the predefined time interval.
- 57.** The system of claim 55, wherein the plurality of relative component values comprises: a first relative component value for each term based on a first portion of the active period of the predefined time interval; a second relative component value for each term based on a second portion of the active period of the predefined time interval; and a third relative component value for each term based on a third portion of the active period of the predefined time interval.
- 58.** The system of claim 55, wherein the at least one processor is further configured to determine a set of terms further comprises reducing a size of a term set by applying at least one filter.
- 59.** The system of claim 58, wherein the at the least one filter includes at least one of:
eliminating candidate terms that are present in a controlled vocabulary;
applying a general fuzzy matching algorithm;
applying a folding routine; and
eliminating candidate terms comprising a stopword.
- 60.** The system of claim 55, wherein the predetermined criteria comprises at least one of: at least one term instance in a minimum number of records;
at least one term instance in at least one record for a minimum number of record periods;

- at least one term instance in a maximum percentage of records from a predefined first portion of the time interval of the data set;
- a ratio of a number of records with at least one term instance from a predefined first portion of the time interval of the data set to a number of records with at least one term instance of the same term from a predefined second portion of the time interval of the data set of at least about 1:2; and
- at least one term instance in a plurality of records having different author data.
- 61.** The system of claim 55, wherein the at least one processor is further configured to determine at least one relative component value comprises assigning a numeric value for each relative component value based on at least one trend analysis for the term.
- 62.** The system of claim 61, wherein the trend analysis comprises trend analysis of time series data.
- 63.** The system of claim 55, wherein the at least one processor is further configured to determine at least one relative component value comprises assigning a numeric value for each relative component value based on at least one of the following:
- publication activity of a last portion of an active period minus publication activity of a first portion of the active period;
 - publication activity of a first recent portion of the active period minus publication activity of an immediate prior portion of the active period;
 - slope from a midpoint portion of the active period to a second recent portion of the active period; and
 - slope from a first point portion of the active period to the mid-point portion of the active period.
- 64.** The system of claim 55, wherein the at least one processor is further configured to assign a numeric value for each relative component value based on at least one trend analysis for the term.
- 65.** The system of claim 64, wherein the trend analysis comprises trend analysis of time series data.
- 66.** The system of claim 55, wherein the at least one processor is further configured to apply a predetermined weight to each relative component value to determine the numeric score.

67. The system of claim 55, wherein the at least one processor is further configured to apply a predetermined weight to each relative component value; and aggregating the relative component values.
68. The system of claim 55, wherein the at least one processor is further configured to generate at least one indicator for a set of records sharing at least one feature, the at least one indicator being based on at least one of:
- summation of scores of terms having scores meeting a predetermined threshold;
 - number of instances of terms having scores meeting a predetermined threshold;
 - average score of terms having scores meeting a predetermined threshold for set of records;
 - summation of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - average of a natural logarithm of scores of terms having scores meeting a predetermined threshold;
 - summation of a square root of scores of terms having scores meeting a predetermined threshold;
 - average of a square root of scores of terms having scores meeting a predetermined threshold;
 - number of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;
 - percentage of records from the set containing at least one instance of terms having scores meeting a predetermined threshold;
 - average score of terms having scores meeting a predetermined threshold and appearing within a record;
 - percentage of records within the set meeting a predetermined average score;
 - percentage of records within the set having a predetermined number of instances of terms having scores meeting a predetermined threshold and appearing within a record; and
 - number of unique terms having scores meeting a predetermined threshold;
69. The system of claim 68, wherein the at least one processor is further configured to assign a value to each indicator.

- 70. The system of claim 68, wherein the at least one processor is further configured to calculate a numeric summation of the square root of scores of terms having scores meeting a predetermined threshold for each first instance of the term within each record of the set.
- 71. The system of claim 70, wherein the at least one processor is further configured to normalize the numeric summation based on the number of records in the set, or square root of the number of records in the set.
- 72. The system of claim 68, wherein the feature relates to information associated with at least one of the following: a person; an organization; a country; a data source; a document type; a document section; a technology; a technical domain; a technical field; a time interval; a geographical location; and a patent class.
- 73. The system of claim 55, wherein the at least one processor is further configured to display, the emergent terms comprising the numeric scores.
- 74. The system of claim 68, wherein the at least one processor is further configured to display the at least one indicator.
- 75. The system of claim 74, the at least one processor is further configured to display a plurality of indicators for a plurality of record sets.
- 76. The method of claim 1, medium of claim 36, or system of claim 55, wherein the score is utilized in a Competitive Technical Intelligence (CTI) application, or a Technology Readiness Assessment (TRA), or a combination thereof.