CLAIMS:

1. A method comprising:

(a) receiving data from a first data source;

(b) performing a source analysis process that includes:

attributing said first data source at a context level, a source file level, and a content level, in accordance with rules, thus yielding an attribute of said first data source;

analyzing said data to identify a characteristic of said data that confounds a meaning of said data, thus yielding a confounding characteristic;

calculating a qualitative measure of said attribute of said first data source, thus yielding a weighted attribute of said first data source;

calculating a qualitative measure of said confounding characteristic, thus yielding a weighted confounding characteristic; and

analyzing said weighted attribute of said first data source and said weighted confounding characteristic, to produce a disposition that includes disposition instructions;

(c) processing said data in accordance with said disposition instructions, thus yielding extracted data;

(d) transmitting said extracted data to a downstream process;

(e) determining, based on said weighted attribute of said first data source and said weighted confounding characteristic, to use said first data source as a seed of an automated data source discovery process;

(f) executing said automated data discovery process using search terms based on content of said first data source to discover a second data source; and

(g) performing said source analysis process on data from said second data source.

2. The method of claim 1, further comprising:

generating feedback based on said disposition; and

improving said method based on said feedback.

3. The method of claim 1, wherein said analyzing is performed in a dimension selected from the group consisting of entity extraction, semantic disambiguation, sentiment analysis, language extraction, linguistic transformation, and basic metadata.

4. The method of claim 1, wherein said confounding characteristic is selected from the group consisting of sarcasm, neologism, grammar variation, inappropriately phrased text, punctuation, polylingual data, spelling, obfuscation, encryption, context, and use of a combination of media.

5. The method of claim 1, wherein said disposition is selected from the group consisting of:

(a) set a rule that files analogous to said first data source be taken in and stored in toto,

(b) divide up files from said first data source and take in and store only parts that meet certain criteria,

(c) take in and store an entire file from said first data source, but flag data with a source-specific quality level indicator,

(d) set a rule that files from said first data source always be rejected, and

(e) tentatively take in and store files from said first data source, but hold them pending additional corroboration.

6. A system comprising:

a processor; and

a memory that contains instructions that are readable by said processor to cause said processor to:

(a) receive data from a first data source;

(b) perform a source analysis process in which said instructions cause said processor to:

attribute said first data source at a context level, a source file level, and a content level, in accordance with rules, thus yielding an attribute of said first data source;

analyze said data to identify a characteristic of said data that confounds a meaning of said data, thus yielding a confounding characteristic;

calculate a qualitative measure of said attribute of said first data source, thus yielding a weighted attribute of said first data source;

calculate a qualitative measure of said confounding characteristic, thus yielding a weighted confounding characteristic; and

analyze said weighted attribute of said first data source and said weighted confounding characteristic, to produce a disposition that includes disposition instructions;

(c) process said data in accordance with said disposition instructions, thus yielding extracted data;

(d) transmit said extracted data to a downstream process;

(e) determine, based on said weighted attribute of said first data source and said weighted confounding characteristic, to use said first data source as a seed of an automated data source discovery process;

(f) execute said automated data discovery process using search terms based on content of said first data source to discover a second data source; and

(g) perform said source analysis process on data from said second data source.

7. The system of claim 6, wherein said instructions also cause said processor to:

generate feedback based on said disposition; and

improve said method based on said feedback.

8. The system of claim 6, wherein said instructions that cause said processor to analyze said data cause said processor to analyze said data in a dimension selected from the group consisting of entity extraction, semantic disambiguation, sentiment analysis, language extraction, linguistic transformation, and basic metadata.

9. The system of claim 6, wherein said confounding characteristic is selected from the group consisting of sarcasm, neologism, grammar variation, inappropriately phrased text,

punctuation, polylingual data, spelling, obfuscation, encryption, context, and use of a combination of media.

10. The system of claim 6, wherein said disposition is selected from the group consisting of:

(a) set a rule that files analogous to said first data source be taken in and stored in toto,

(b) divide up files from said first data source and take in and store only parts that meet certain criteria,

(c) take in and store an entire file from said first data source, but flag data with a source-specific quality level indicator,

(d) set a rule that files from said first data source always be rejected, and

(e) tentatively take in and store files from said first data source, but hold them pending additional corroboration.

11. A non-transitory computer readable storage medium, comprising instructions that are readable by a processor to cause said processor to:

(a) receive data from a first data source;

(b) perform a source analysis process in which said instructions cause said processor to:

attribute said first data source at a context level, a source file level, and a content level, in accordance with rules, thus yielding an attribute of said first data source;

analyze said data to identify a characteristic of said data that confounds a meaning of said data, thus yielding a confounding characteristic;

calculate a qualitative measure of said attribute of said first data source, thus yielding a weighted attribute of said first data source;

calculate a qualitative measure of said confounding characteristic, thus yielding a weighted confounding characteristic; and

analyze said weighted attribute of said first data source and said weighted confounding characteristic, to produce a disposition that includes disposition instructions;

(c) process said data in accordance with said disposition instructions, thus yielding extracted data;

(d) transmit said extracted data to a downstream process;

(e) determine, based on said weighted attribute of said first data source and said weighted confounding characteristic, to use said first data source as a seed of an automated data source discovery process;

(f) execute said automated data discovery process using search terms based on content of said first data source to discover a second data source; and

(g) perform said source analysis process on data from said second data source.

12. The non-transitory computer readable storage medium of claim 11, wherein said instructions also cause said processor to:

generate feedback based on said disposition; and

improve said method based on said feedback.

13. The non-transitory computer readable storage medium of claim 11, wherein said instructions that cause said processor to analyze said data cause said processor to analyze said data in a dimension selected from the group consisting of entity extraction, semantic disambiguation, sentiment analysis, language extraction, linguistic transformation, and basic metadata.

14. The non-transitory computer readable storage medium of claim 11, wherein said confounding characteristic is selected from the group consisting of sarcasm, neologism, grammar variation, inappropriately phrased text, punctuation, polylingual data, spelling, obfuscation, encryption, context, and use of a combination of media.

15. The non-transitory computer readable storage medium of claim 11, wherein said disposition is selected from the group consisting of:

(a) set a rule that files analogous to said first data source be taken in and stored in toto,

(b) divide up files from said first data source and take in and store only parts that meet certain criteria,

(c) take in and store an entire file from said first data source, but flag data with a source-specific quality level indicator,

(d) set a rule that files from said first data source always be rejected, and

(e) tentatively take in and store files from said first data source, but hold them pending additional corroboration.