CLAIMS

What is claimed is:

1. A computer-implemented method comprising:

prior to a bid time:

receiving, by a computer, a first set of context terms received from a client device;

applying, by the computer, a machine-learning model on the first set of context terms to output a set of beacon terms extracted from a corpus database, wherein the computer is configured according to a plurality of corpus terms stored in the corpus database to determine a plurality of co-occurrence probabilities corresponding to the plurality of corpus terms, and wherein each beacon term corresponds to a corpus term that satisfies a threshold co-occurrence probability;

calculating, by the computer, a plurality of page scores for a plurality of corpus webpages stored in the corpus database based upon the set of beacon terms and the first set of context terms;

identifying, by the computer, in the plurality of corpus webpages a set of contextual webpages having page scores satisfying a threshold;

generating, by the computer, according to the first set of context terms and a second set of context terms received from the client device an updated set of beacon terms;

calculating, by the computer, one or more updated page scores for one or more corpus webpages stored in the corpus database based upon the updated set of beacon terms, the first set of context terms, and the second set of context terms;

updating, by the computer, the set of contextual webpages based upon the one or more updated page scores; and

storing, by the computer into a campaign database, campaign data of a user comprising the set of beacon terms and the set of contextual webpages, the campaign data configured for executing at a future bid time a real-time bidding selection operation for one or more available webpages during the real-time bidding selection operation.

2. The method according to claim 1, further comprising:

extracting, by the computer, one or more embeddings for the first set of context terms, wherein the set of beacon terms comprises one or more corpus terms in the corpus database having

a feature vector satisfying a threshold distance from the one or more embeddings for the first set of context terms.

3. The method according to claim 2, further comprising:

calculating, by the computer, a context score for each context term in the first set of context terms is based upon at least one of: a word probability, a word co-occurrence count, and a word-to-word probabilistic implication, wherein each embedding of each context term for identifying the one or more beacon terms is based upon the context score for the context term.

4. The method according to claim 3, further comprising:

setting, by the computer, each context score as a default score for each context term; and

adjusting, by the computer, the context score for the context term in accordance with one or more user inputs.

5. The method according to claim 1, wherein the first set of context terms comprises at least one out-of-context term.

6. The method according to claim 1, further comprising:

generating, by the computer, context data for the set of contextual webpages based upon one or more user inputs received from the client device; and

transmitting, by the computer, the context data to the client device for display at a graphical user interface (GUI) of the client device.

7. The method according to claim 1, wherein the computer receives each context term from the client device via a query configuration webpage.

8. The method according to claim 2, further comprising:

requesting, by the computer, webpage data of a webpage comprising page content and metadata according to a uniform resource locator (URL); and

storing, by the computer, the webpage data into the corpus database as a preprocessed page, the preprocessed page stored into a data record for the preprocessed page, the data record comprising at least a portion of the page content and the metadata, a timestamp, the URL, and one or more content tags.

9. The method according to claim 1, further comprising:

receiving, by the computer, from a bid server an availability list of one or more available webpages requesting bids from a bid system; and

calculating, by the computer, a real-time page score for each available webpage in the availability list based, at least in part, upon a number of occurrences of one or more campaign terms comprising the set of updated beacon terms and one or more in-context terms in the available webpage.

10. The method according to claim 9, further comprising:

identifying, by the computer, a bidding-list of webpages comprising each of the available webpages of the availability list satisfying a bid threshold.

11. A system comprising:

a corpus database comprising non-transitory storage medium configured to store at least a portion of a plurality of Internet content data corresponding to a uniform resource locator (URL) and page context score;

a campaign database comprising non-transitory storage medium configured to store campaign data for a plurality of users, the campaign data configured for executing a real-time bidding selection operation for one or more available webpages; and

a server comprising a processor configured to:

prior to a bid time:

receive a first set of context terms received from a client device;

apply a machine-learning model on the first set of context terms to output a set of beacon terms extracted from the corpus database, wherein the machine-learning model is configured according to a plurality of corpus terms stored in the corpus database to determine a plurality of co-occurrence probabilities corresponding to the plurality of corpus terms, and wherein each beacon term corresponds to a corpus term that satisfies a threshold co-occurrence probability;

calculate a plurality of page scores for a plurality of corpus webpages stored in the corpus database based upon the set of beacon terms and the first set of context terms;

identify in the plurality of corpus webpages a set of contextual webpages having page scores satisfying a threshold;

generate according to the first set of context terms and a second set of context terms received from the client device to output an updated set of beacon terms;

calculate one or more updated page scores for one or more corpus webpages stored in the corpus database based upon the updated set of beacon terms, the first set of context terms, and the second set of context terms;

update the set of contextual webpages based upon the one or more updated page scores; and

store, into the campaign database, the campaign data of a user comprising the set of beacon terms and the set of contextual webpages, the campaign data configured for executing at a future bid time the real-time bidding selection operation for the one or more available webpages during the real-time bidding selection operation.

12. The system according to claim 11, wherein the server is further configured to:

generate context data for the set of contextual webpages based upon one or more user inputs received from the client device; and

transmit the context data to the client device for display at a graphical user interface (GUI) of the client device.

13. The system according to claim 11, wherein the server is further configured to:

extract one or more embeddings for the first set of context terms, wherein the set of beacon terms comprises one or more corpus terms in the corpus database having a feature vector satisfying a threshold distance from the one or more embeddings for the first set of context terms.

14. The system according to claim 11, wherein the server is further configured to:

receive from a bid server an availability list of one or more available webpages requesting bids from a bid system; and

calculate a real-time page score for each available webpage in the availability list based, at least in part, upon a number of occurrences of one or more campaign terms comprising the set of updated beacon terms and one or more in-context terms in the available webpage.

15. A computer readable medium containing machine-executable program instructions, wherein execution of the program instructions by one or more processors of a computer system causes the one or more processors to execute the steps of:

prior to a bid time:

receiving a first set of context terms received from a client device;

applying a machine-learning model on the first set of context terms to output a set of beacon terms extracted from a corpus database, wherein the computer is configured according to a plurality of corpus terms stored in the corpus database to determine a plurality of co-occurrence probabilities corresponding to the plurality of corpus terms, and wherein each beacon term corresponds to a corpus term that satisfies a threshold co-occurrence probability;

calculating a plurality of page scores for a plurality of corpus webpages stored in the corpus database based upon the set of beacon terms and the first set of context terms;

identifying in the plurality of corpus webpages a set of contextual webpages having page scores satisfying a threshold;

generating according to the first set of context terms and a second set of context terms received from the client device an updated set of beacon terms;

calculating one or more updated page scores for one or more corpus webpages stored in the corpus database based upon the updated set of beacon terms, the first set of context terms, and the second set of context terms;

updating the set of contextual webpages based upon the one or more updated page scores; and

storing, into a campaign database, campaign data of a user comprising the set of beacon terms and the set of contextual webpages, the campaign data configured for executing at a future bid time a real-time bidding selection operation for one or more available webpages during the real-time bidding selection operation.

16. The computer readable medium of claim 15, wherein the one or more processors further executing the steps of:

extracting one or more embeddings for the first set of context terms, wherein the set of beacon terms comprises one or more corpus terms in the corpus database having a feature vector satisfying a threshold distance from the one or more embeddings for the first set of context terms.

17. The computer readable medium of claim 16, wherein the one or more processors further executing the steps of:

calculating a context score for each context term in the first set of context terms is based upon at least one of: a word probability, a word co-occurrence count, and a word-to-word probabilistic implication, wherein each embedding of each context term for identifying the one or more beacon terms is based upon the context score for the context term.

18. The computer readable medium of claim 17, wherein the one or more processors further executing the steps of:

setting each context score as a default score for each context term; and

adjusting the context score for the context term in accordance with one or more user inputs.

19. The computer readable medium of claim 15, wherein the one or more processors further executing the steps of:

generating context data for the set of contextual webpages based upon one or more user inputs received from the client device; and

transmitting the context data to the client device for display at a graphical user interface (GUI) of the client device.

20. The computer readable medium of claim 15, wherein the one or more processors further executing the steps of:

receiving, by the computer, from a bid server an availability list of one or more available webpages requesting bids from a bid system; and

calculating, by the computer, a real-time page score for each available webpage in the availability list based, at least in part, upon a number of occurrences of one or more campaign terms comprising the set of updated beacon terms and one or more in-context terms in the available webpage.