



(51) International Patent Classification:  
**G06F 7/00** (2006.01)

(21) International Application Number:  
PCT/US2014/049292

(22) International Filing Date:  
31 July 2014 (31.07.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/860,874 31 July 2013 (31.07.2013) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **SILVA, Alcino, J.**; 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US). **LANDRETH, Anthony**; 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US). **BICKLE, John**; 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US).

(74) Agent: **HILL, James W., MD**; McDermott Will & Emery LLP, 4 Park Plaza, Suite 1700, Irvine, California 92614 (US).

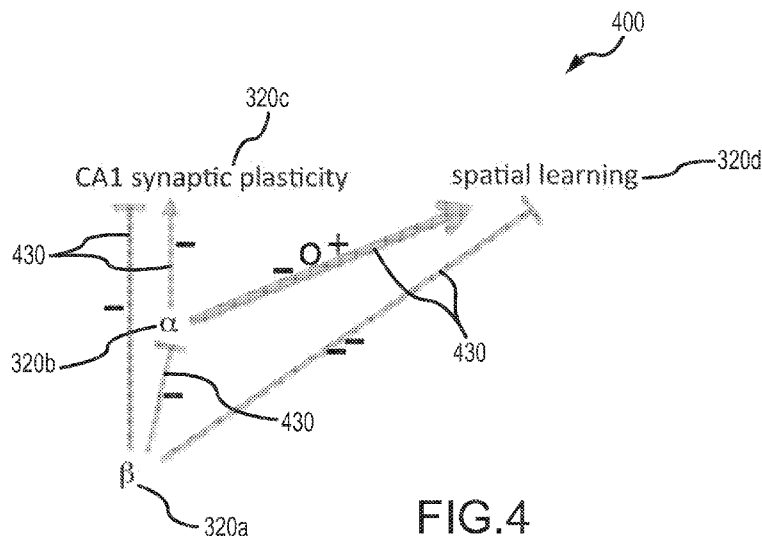
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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, SA, 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, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: APP TO BUILD MAPS OF RESEARCH FINDINGS



(57) Abstract: Research data maps can be generated from a database, based on a user selection of a parameter representing a selected agent or target of an experiment, the database containing experimental results data corresponding to relationships, determined by experiments, between the agent or target and other agents and/or targets. The research map includes nodes and edges, each of the nodes representing an agent and/or target and connected to another node by one of the edges; wherein each of the edges indicates a causal relationship between connected nodes based on one or more of the determined relationships. A hypothesis may be generated positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or target.

## APP TO BUILD MAPS OF RESEARCH FINDINGS

### Related Applications

[0001] The present application claims the benefit of priority from U.S. Provisional Patent Application Serial No. 61/860,874 entitled, "APP TO BUILD MAPS OF RESEARCH FINDINGS," filed on July 31, 2013, which is hereby incorporated by reference in its entirety for all purposes.

### Background

[0002] The growth of the scientific literature in the last 30 years has been astronomic. The library of medicine now includes more than 20 million articles, and one discipline (neuroscience) includes nearly two million research articles. Simplified abstractions of published information could be used to characterize what is known and to guide research decisions.

### Summary

[0003] The subject technology is illustrated, for example, according to various aspects described below.

[0004] There is a great need to develop maps (simplified abstractions) of published information that could be used to characterize what is known and to guide research decisions. Disclosed herein is a strategy to build maps of research findings. The maps provide a formal language to represent research findings. Disclosed is a taxonomy for experiments and a set of algorithms to generate these maps. This taxonomy allows classification of experiments in a field into a small set of distinct categories, a critical first step in the development of simplified abstractions of research findings. The algorithms represent these experiments in weighted causal networks (i.e., maps).

[0005] Also disclosed is an application that partially automates the process of generating these maps based on research data entered by the user. Beyond helping scientists assess the weight of evidence behind key findings in any given area of research, these maps will also be used to make decisions about what drugs to develop next for a particular medical condition since they represent the relative strength of evidence for different alternatives. They will also be used to determine objectively the contributions of research papers, to evaluate the potential of grant proposals, to better represent information in science, etc. It is also possible

to weight the strength of evidence to be used for purposes outside of science. The complexity of the world, and the immensity of data of various quality in the WWW requires strategies that weight the evidence for the information represented in internet searches. The subject technology could be used to develop a fundamental new type of search engine that gives users responses to their queries that are based on the cumulative evidence scanned from many web pages according to the weighted causal rules outlined in our algorithms.

[0006] Current searches for scientific findings depend on abstracted material that can be obtained from the library of medicine and other similar government sponsored compilations of scientific articles. When searching for scientific items, the user simply gets a long list of abstracts. According to embodiments of the subject technology, it is possible to obtain causal networks that summarize large amounts of information in an easy to use and interactive manner. The application allows users to obtain summaries of papers they have reviewed and entered into the system. These summaries are in the form of causal networks where each connection in the network is weighted to the evidence supporting it. This tool will be invaluable in determining exactly what is known about a subject and how to best contribute to that topic.

[0007] The implementation of exemplary embodiments takes place in at least two phases: in the first phase, users will take advantage of our application to cooperatively generate weighted causal networks of the manuscripts that they have added to the database at the heart of our app. In the second stage of an exemplary implementation, natural language algorithms, field-defined ontologies and machine learning approaches will be used to automate the process of entering data into the database at the heart of the generation of our weighted causal networks. A query-based interactive system will show the user just as much complexity as the user requests, out of all of the data entered into the system. This system will be available to users that want to interact with the data that they entered into it. Users may then see research maps derived from other data in the system, whether it was entered manually or automatically as described above. Organizations, such as pharmaceutical companies, may acquire specific research maps relevant to their ongoing efforts, such as drug development. For example, causal maps would be useful in making decisions about which specific drug target to pursue.

**[0008]** Research data maps can be generated from a database, based on a user selection of a parameter representing a selected agent or target of an experiment, the database containing experimental results data corresponding to relationships, determined by experiments, between the agent or target and other agents and/or targets. The research map includes nodes and edges, each of the nodes representing an agent and/or target and connected to another node by one of the edges; wherein each of the edges indicates a causal relationship between connected nodes based on one or more of the determined relationships. A hypothesis may be generated positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or target.

**[0009]** According to some embodiments, disclosed is a method, comprising: receiving, from a user, a selection of a plurality of parameters; from a database and by a processor, generating a research map providing an indication of a causal relationship between two of the parameters; and displaying the research map on a display device.

**[0010]** The indication of the causal relationship can comprise strength of the causal relationship. The indication of the causal relationship can comprise a positive or negative correlation between the two of the parameters. The database can comprise a record of each of a plurality of experimental findings described in corresponding publications, wherein the record can comprise the causal relationship between the two of the parameters. The method can further comprise generating a hypothesis relating to an additional causal relationship between another two of the parameters; and displaying the hypothesis on the display device.

**[0011]** According to some embodiments, disclosed is a computer implementation system, comprising: an input module that, by a processor, receives a selection of a plurality of parameters from a user; a database; a map generating module that, from the database and by the processor, generates a research map providing an indication of a causal relationship between two of the parameters; and an output module that, by the processor, displays the research map on a display device.

**[0012]** The indication of the causal relationship can comprise strength of the causal relationship. The indication of the causal relationship can comprise a positive or negative correlation between the two of the parameters. The database can comprise a record

of each of a plurality of experimental findings described in corresponding publications, wherein the record can comprise the causal relationship between the two of the parameters.

**[0013]** According to some embodiments, disclosed is a machine-readable medium comprising machine-readable instructions for causing a processor to execute a method comprising, comprising: receiving, from a user, a selection of a plurality of parameters; from a database and by a processor, generating a research map providing an indication of a causal relationship between two of the parameters; and displaying the research map on a display device.

**[0014]** The indication of the causal relationship can comprise strength of the causal relationship. The indication of the causal relationship can comprise a positive or negative correlation between the two of the parameters. The database can comprise a record of each of a plurality of experimental findings described in corresponding publications, wherein the record can comprise the causal relationship between the two of the parameters.

**[0015]** According to some embodiments, disclosed is a method comprising: receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment; by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target; wherein the research map can comprise nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges; wherein one of the nodes represents the selected agent or the selected target; wherein each of the edges can comprise an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and outputting, by a processor, a hypothesis positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

**[0016]** The method can further comprise: generating a hypothesis map based on the hypothesis. The method can further comprise: displaying a graphical representation of the hypothesis map on a display device. The method can further comprise: displaying a graphical representation of the research map on a display device. The indication of the causal relationship can comprise a strength of the causal relationship. The indication of the causal

relationship can comprise a positive or negative correlation between the connected nodes. The indication of the causal relationship can comprise an indication of a type of the determined relationship on which the causal relationship is based. The experimental results data can comprise a record of each of a plurality of experimental results described in corresponding publications. The experimental results data can correspond to results indicators relating to categories of the at least one performed experiment, the categories selected from the group consisting of positive manipulation, negative manipulation, non-intervention, and mediation. The method can further comprise: receiving, from a user, additional experimental results data of an experiment, the experimental results data comprising indicators of (i) two variables, (ii) an experimental relationship indicator corresponding to two variables, and (iii) a result indicator corresponding to a type of the experiment.

[0017] According to some embodiments, disclosed is a computer implementation system, comprising: an input module that, by a processor, receives a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment; a database containing experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target; a map generating module that, from the database and by the processor, generates a research map based on the selection; wherein the research map can comprise nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges; wherein one of the nodes represents the selected agent or the selected target; wherein each of the edges can comprise an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and a hypothesis module that posits a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

[0018] The computer implementation system can further comprise: a display module that displays a graphical representation of the research map. The indication of the causal relationship can comprise a strength of the causal relationship. The indication of the causal relationship can comprise a positive or negative correlation between the connected nodes. The indication of the causal relationship can comprise an indication of a type of the determined relationship on which the causal relationship is based. The experimental results data can comprise a record of each of a plurality of experimental results described in

corresponding publications. The experimental results data can correspond to results indicators relating to categories of the at least one performed experiment, the categories selected from the group consisting of positive manipulation, negative manipulation, non-intervention and mediation.

**[0019]** According to some embodiments, disclosed is a machine-readable medium comprising machine-readable instructions for causing a processor to execute a method comprising, comprising: receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment; by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target; wherein the research map can comprise nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges; wherein one of the nodes represents the selected agent or the selected target; wherein each of the edges can comprise an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and outputting, by a processor, a hypothesis positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

**[0020]** According to some embodiments, disclosed is a method comprising: receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment; by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target; generating a research map, the research map comprising nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges; wherein one of the nodes represents the selected agent or the selected target; wherein each of the edges can comprise an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and outputting, by a processor, a hypothesis map positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

[0021] Additional features and advantages of the subject technology will be set forth in the description below, and in part will be apparent from the description, or may be learned by practice of the subject technology. The advantages of the subject technology will be realized and attained by the structure particularly pointed out in the written description and claims hereof as well as the appended drawings.

[0022] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory and are intended to provide further explanation of the subject technology as claimed.

#### Brief Description of the Drawings

[0023] The accompanying drawings, which are included to provide further understanding of the subject technology and are incorporated in and constitute a part of this specification, illustrate aspects of the subject technology and together with the description serve to explain the principles of the subject technology.

[0024] FIG. 1 shows a flow chart with steps involved in building a research map, according to some embodiments of the subject technology.

[0025] FIG. 2 shows exemplary experiments and findings of operation 110 in FIG. 1, according to some embodiments of the subject technology.

[0026] FIG. 3 shows a table of individual experiments of operation 120 in FIG. 1, according to some embodiments of the subject technology.

[0027] FIG. 4 shows an exemplary research map of operation 130 in FIG. 1, according to some embodiments of the subject technology.

[0028] FIG. 5 shows an exemplary hypothesis of operation 140 in FIG. 1, according to some embodiments of the subject technology.

[0029] FIG. 6 shows a research map representing results in an exemplary paper, according to some embodiments of the subject technology.

[0030] FIG. 7 shows a flow chart with steps involved in calibrating a score for a strength of an experimental connection, according to some embodiments of the subject technology.



[0031] FIG. 8 shows an exemplary interface for interacting with a research map, according to some embodiments of the subject technology.

[0032] FIG. 9 shows an exemplary interface for interacting with a research map, according to some embodiments of the subject technology.

[0033] FIG. 10 shows an exemplary interface for interacting with a research map, according to some embodiments of the subject technology.

[0034] FIG. 11 shows an exemplary system, according to some embodiments of the subject technology.

[0035] FIG. 12 shows an exemplary system, according to some embodiments of the subject technology.

[0036] FIG. 13 shows an exemplary system, according to some embodiments of the subject technology.

#### Detailed Description

[0037] In the following detailed description, numerous specific details are set forth to provide a full understanding of the subject technology. It will be apparent, however, to one ordinarily skilled in the art that the subject technology may be practiced without some of these specific details. In other instances, well-known structures and techniques have not been shown in detail so as not to obscure the subject technology.

[0038] The increasing volume, complexity and interconnectedness of published studies in neuroscience make it difficult to determine what is known, what is uncertain, and how to contribute effectively to one's field. There is a pressing need to develop automated strategies to help researchers navigate the vastness of the published record. Simplified, interactive and unbiased representations of previous findings (i.e. research maps) would be invaluable in preparing research surveys, in guiding experiment planning, and in evaluating research plans and contributions. Principles used in weighing research findings, including reproducibility and convergence, can be applied to build research maps. Discussed herein are systematic, comprehensive, interactive, and user-friendly research maps. These maps could revolutionize the way we review the literature, plan experiments, fund and publish science.

**[0039]** Published research in neuroscience has grown massive. The past three decades have accumulated more than 1.6 million articles alone. The rapid expansion of the published record has been accompanied by an unprecedented widening of the range of concepts, approaches and techniques that individual neuroscientists are expected to be familiar with. The cutting edge of neuroscience is increasingly defined by studies demanding researchers in one area (e.g., molecular and cellular neuroscience) to have more than a passing familiarity with the tools, concepts and literature of other areas (e.g., systems or behavioral neuroscience). As research relevant to a topic expands, it becomes increasingly more likely that researchers will be either overwhelmed or unaware of relevant results (or both). Consequently, there is a pressing need for new tools to help neuroscientists navigate the complexity and size of published information. There is an urgent need to develop research maps -simplified, interactive and unbiased representations of research findings- not only to clarify what has been accomplished, but also to serve as guides in choosing what will be accomplished next.

**[0040]** Currently, the Library of Medicine has nearly 25 million articles in its records. These include an estimated 250 million experiments, with nearly 5 million added every year. It is increasingly difficult for biomedical researchers to be aware of even a small fraction of this ever-growing body of research. Additionally, the biomedical research landscape has become more interconnected; Biologists are increasingly required to master concepts and information in several different fields, some of which they are only passingly familiar. The challenges are not limited to the sheer number of experiments and publications that biologists are required to remember and understand in detail. The greatest challenge is that biologists are asked to understand the implications of this body of complex experimental relations. Therefore, there is a great need for approaches that facilitate the process of tracking and integrating information across different research papers often coming from very different disciplines in biology. A tool capable of generating causal networks of available information from different fields and specialties in the biological sciences would be enormously useful in integrating published information, planning future experiments and devising new research directions. Such a tool would also be extremely useful for objectively estimating the possible impact of research proposals and judging the contributions of experimental findings.

[0041] Review papers and opinion articles are a traditional source of general summaries of research concepts. However, there are drawbacks associated with these traditional summaries. First, they are not dynamically updated, and usually reflect the state of a field at least one to two years old. In fields with fast research growth, such as the health sciences, with an estimated 13,000 experiments published every day, this is a significant factor. Another drawback is that traditional reviews reflect the research biases of the writer. The subject technology uses the framework and integration principles discussed below to generate interactive graphical representations of experiments (Research Maps).

[0042] Mapping relevant research (i.e., determining the information directly relevant to a particular research topic) is closely related to experiment planning (i.e., conceiving and evaluating a potential series of future experiments). In choosing which experiment to perform next, we advance with the hope that our knowledge and training will provide firm footing for a trek into unknown territory. But without research maps, we risk missing key information while planning new experiments. We also risk conducting redundant experiments.

[0043] Research maps can be developed in three different ways. First, databases of unambiguous and concise representations of experiments and their results can be built. Second, to assess the evidential weight in favor of hypotheses found among these representations, familiar kinds of reasoning used in our respective fields can be automated to evaluate evidence. For example, reproducibility and convergence of research findings are two of the principles universally used in neuroscience to weigh research findings. Reproducibility is the ability of an experimental finding to be replicated independently with identical or similar procedures. Convergence reflects the ability of very different experiments to point to a single conclusion. Quantitative measures of reproducibility and convergence could be used to weight the evidence for embedded causal hypotheses in research maps (FIG. 1). Third, effective protocols for sharing these representations can be developed, so that we can combine knowledge across research communities.

[0044] An important component of a “research map” is a database of research summaries and their results. This database could then be used to generate an interactive graphical summary (i.e. a map) of that research. The flow chart 100 in FIG. 1 illustrates the key steps used to create a research map, including the extraction of experiments and findings

from the primary research literature (operation 110), the derivation of a database of those findings (operation 120), which are then used to derive an integrated graphical representation of those experiments (i.e., a research map; operation 130) and suggest causal hypotheses (operation 140). A user reading a research map would be able to survey a specific research area at different levels of resolution, from coarse summaries of findings (operation 130) to fine-grained accounts of experimental results. The primary function of a research map is to display no more and no less information to a user than is necessary for the researcher's purposes.

[0045] According to some embodiments, as shown in operation 110 of FIG. 1, single experiments from a research publication are extracted. These individual experiments, along with other relevant information about methods and authors, could be captured into a nano-publication.

[0046] According to some embodiments, as shown in operation 120 of FIG. 1, the experiments could then be entered into a database as experimental results data 300 with a format optimized for the extraction of graphic representations of those experiments (see below). The examples listed involve experiments with two variables 320 (e.g., proteins  $\beta$  and  $\alpha$  in experiment 208). The relationship between the two variables 320 can be represented by experimental relationship indicator 330 and recorded in results indicators 350, showing the “-” symbol represents negative alteration experiments, in which the activity or levels of one variable 320 were decreased and measurements were taken on another. The “+” symbol represents positive alteration experiments, where the activity or levels of one variable 320 were increased and measurements were taken on another. The “O” symbol represents non-intervention experiments, which involve no manipulation of either variable 320. Instead, the activity or levels of both variables 320 were measured. The arrowheads in “+” and “-” experiments point away from the manipulated variables 320. In the “O” experiment, the arrowhead points away from the variable 320 whose changes preceded changes in the other variable 320.

[0047] According to some embodiments, as shown in operation 130 of FIG. 1 and research map 400 of FIG. 4, a research map 400 (integrated graphic representation) is derived from the database in operation 120. This research map 400 provides a convenient, although course, visual summary of the results indicators 350 listed in operation 110. The weight of

the resultant relationship indicators 430 (e.g., arrows) between variables 320 (e.g., 320a, 320b, 320c, and 320d) represent the strength of the evidence supporting the proposed causal connection denoted by the experimental relationship indicators 330. For example, the connection with a heavy arrow 430 is supported by the three types of convergent evidence outlined in operation 120, while the other indicators 430 with lighter arrows are supported by weaker evidence.

[0048] According to some embodiments, as shown in operation 140 of FIG. 1 and hypothesis 500, beyond providing objective summaries of experimental findings, research maps 400 can also be used for hypothesis building based on the variables 320 (e.g., 320a-d). For example, proposed relationship indicators 530 can be hypothesized based on the resultant relationship indicators 430 between variables 320 (e.g., 320a-d). For example, shown is a graphic representation of the hypothesis that  $\beta$  inhibits  $\alpha$  and that  $\alpha$  activation is needed for triggering synaptic plasticity in CA1 which in turn is required for spatial learning.

[0049] According to some embodiments, as shown in operation 150 of FIG. 1, with this hypothesis in hand, one could use the research map in operation 130 to choose experiments that could strengthen (or weaken) this hypothesis. The figure lists two experiments (“+”, “O”) that could test the causal connection between  $\alpha$  protein activation and CA1 plasticity. “+” experiment would increase  $\alpha$  protein and look at effects in plasticity, and the “O” would determine whether increases in  $\alpha$  protein precede CA1 synaptic plasticity.

[0050] Primary research articles often contain summaries of prior research and statements concerning the significance of findings presented. Additionally, review articles can help to place specific collections of findings in a broader more integrated perspective. However valuable, the individual perspectives in research papers and review articles are not always objective and balanced. Frequently, they do not reflect all of the relevant information available for the topic being reviewed. Thus, in addition to these personal perspectives, it would be useful to consult exhaustive, inclusive and integrated databases (i.e., research maps) concerning the results and experimental strategies of an area or topic of interest.

[0051] To enhance the accessibility of research maps, each assertion would be stated in an unambiguous vocabulary. There are now numerous such vocabularies for automated reasoning, called ontologies (e.g. available through the National Center for Biomedical Ontologies or NCBO). Unlike natural languages (e.g. English), biomedical

ontologies map one entity into one term. For instance, the word ‘nucleus’ is ambiguous between a cluster of cells, the nucleus of a single cell, and an atomic nucleus. The different senses of ‘nucleus’ receive different terms in biomedical ontologies, so that when data is annotated with one of these terms, there is no ambiguity to confound a search over that data, and no ambiguity to confound automated reasoning.

**[0052]** To date, the most extensive effort toward developing an ontology for neuroscience has been undertaken by the Neuroscience Information Framework (NIF). The NIF has collected a dynamic lexicon of over 19,000 neuroscience terms to describe neural structures and functions. The lexicon is built from the NIF standard ontologies (NIFSTD) (Larson and Martone, 2009). To make these vocabularies available to non-specialists, the NIF group has built a web app, NeuroLex, from which a user can easily find the right terms to describe a phenomenon or protocol.

**[0053]** Ontologies like the NIFSTD provide materials for composing unambiguous representations of neuroscience research in a format sometimes called “nano-publication” (Groth, Gibson, and Velterop, 2010). A nanopublication is the smallest unit of publishable information that can be uniquely identified and attributed to its author(s). Each of the eight experiments 201-208 in FIG. 2 could be reported in a single conventional research paper, or in eight nano-publications. Nano-publications usually include a subject-predicate-object structure, e.g. gene-alpha (subject) is linked to (predicate) protein-beta (object). Nano-publications also provide meta-data concerning, for example, the experimental methods used, as well as information about the authors (cf. <http://nanopub.org>). Together, these components of a nano-publication tell us no more than what we need to know, when we search for specific results in the published literature.

**[0054]** Nano-publications are a promising basis for building research maps. But to determine the evidential standing of the assertions found in nano-publications, it is key to know how and whether those assertions fit together: for example, whether the findings underlying those assertions reproducible or whether there exist different sets of experiments converging on similar conclusions. Informally asking these questions while conducting a literature review facilitates development of an intuitive sense of the robustness of a result or finding. With that sense, a decision is made whether to trust a hypothesis enough to plan future related experiments. To be useful, causal connections represented in research maps

would be weighted according to principles, including reproducibility and convergence, that neuroscientists use to weigh evidence for findings in their respective fields. For example, in FIG. 2 an experiment set 200 includes experiments 201, 202, 203, 204, 205, 206, 207, and 208 for three fundamentally different types of experiments supporting the idea that protein  $\alpha$  is involved in spatial learning, while there is less experimental support for other potential causal connections listed in that figure. Neuroscientists have greater confidence in findings when they converge across different kinds of experiments. Similarly, results reproduced by multiple related experiments are deemed more reliable. For example, experiments 207 and 208 in FIG. 2 both resulted in decreases in the activity of protein  $\alpha$  despite different methods to disrupt protein  $\beta$  (pharmacology and genetics). Reproducibility and convergence could be used to weight the evidence represented in research maps, and this would help identify strong versus weak results. To accomplish this, however, the experiments are organized into categories.

[0055] For example, some neuroscience experiments are designed to decrease the probability of an event's occurrence, such as an inhibitory drug administered to prevent a receptor's action, or a lesion induced to impair a brain region's function (e.g., experiments 201, 202, 205, 206, 207, and 208 in FIG. 2). Such experiments help us to determine the necessity of a specific phenomenon for the occurrence of another. Other experiments are designed to trigger an event, such as the expression of a gene, or the activation of a brain region (e.g., experiment 204 in FIG. 2). These experiments inform us as to the sufficiency of an event relative to the occurrence of another. In another common type of neuroscience experiment, no variable is intentionally manipulated and the goal is simply to describe how two phenomena co-vary, such as the activity of two molecules or two brain regions (e.g., experiment 203 in FIG. 2). Not surprisingly, when the results of these three very different types of experiments agree, neuroscientists usually place more weight on the underlying hypotheses than when the support is incomplete (based on one type of experiment) or when there are contradictions in the results. One could imagine codifying this process in research maps, so that at a glance we could see the connections in research maps with weak and strong evidence. For example, the connection with a heavy arrow in FIG. 4 is supported by the three different kinds of convergent evidence outlined above, while the other connections represented with lighter arrows have weaker evidential support. Unfortunately, it is often difficult to discern from literature searches, involving hundreds of papers and thousands of

experiments, the weight of evidence (degree of convergence and reproducibility) behind any one finding. Research maps could be a solution to this increasingly serious problem.

[0056] In an attempt to represent large bodies of complex information, researchers draw diagrams with arrows (i.e. path diagrams) that stand for causal connections between phenomena, such as interactions between signaling molecules, and neuroanatomical connections (e.g., FIG. 5). These diagrams are useful for organizing existing research and planning future experiments. But, these representations have important limitations. First, they are essentially static representations that do not update as the knowledge base of experimental results changes. Second, these diagrams do not show all of the equally well-supported alternative models that fit the existing data. Third, they do not show the relative weight of the evidence supporting each of the causal connections represented (commonly drawn as arrows). Finally, these diagrams are almost always composed by a small number of authors, and they are rarely systematic or complete. While the corpus of articles contributing to a diagram's composition is explicit in the review's bibliography, that corpus is necessarily subject to sampling biases, since a small number of authors will only be able read so many articles, recall so many facts, and reason over so many variables. Nor is there an attending protocol that could enable others to read the same articles and thereby derive the same diagrams. Research maps could address all of these limitations while keeping many of the features (e.g., simplicity) that make these diagrams attractive to neuroscientists.

[0057] Ideas and strategies from graphical causal modeling (Pearl, 2000; Spirtes et al., 2000) will be useful for generating research maps. For example, very recently, an algorithm was developed that enables a collection of causal models with overlapping variables to be integrated into a unified causal network (cf. Tillman et al., 2009), a critical step in the generation of integrated large-scale causal networks. Imagine, for example the complexities of attempting to integrate many related research maps such as the one in FIG. 4. How could this be accomplished in a systematic and automate manner? The algorithms in graphical causal modeling could help us construct these integrated research maps, and these maps could be dynamically updated as new results emerge in the research record.

[0058] With a dynamic and interactive graphical interface, a scientist could use a research map to survey a field's experimental findings far faster than by reading abstracts or other textual descriptions. Areas with little research investment would be made apparent by



both the sparseness and weakness of connections among their phenomena, enabling researchers to easily identify opportunities to conduct pioneering experiments (for example, the experiments marked by “?” in the table of results indicators 350 in FIG. 3). Currently, contradictions in the literature are difficult to resolve. These contradictions, however, would be accounted for in research maps by weakening the affected causal connections. Additionally, the global perspective afforded by these maps may help neuroscientists identify the source of contradictions or inconsistencies in the experimental record (e.g., by identifying systematic methodological differences between experiments with contradictory results). Research maps may also help address more objectively the quality of the evidence in the research literature. The uneven quality of research contributions is a real problem in science. Research maps will not solve this problem, but because they include databases of the information associated with research findings (e.g., methods, authors, tools, and models used) they may provide strategies to identify systematic problems in the research record.

**[0059]** Research publications normally highlight only a small subset of the research findings described. Most published experiments are not even alluded to in the abstract, and many are relegated to supplemental figures. Sadly, all scientists know that most experiments are not published at all, and lay forgotten in research notebooks. This large body of forgotten research could be reviewed, reported as nanopublications and integrated into research maps. Traditional research papers have to face the limitations of page counts, numbers of allowed figures, the attention span of potential readers, etc. None of these limitations would apply to the nano-publication content of research maps.

**[0060]** Research maps could be constructed as shown in the flow chart in FIG. 1. Training in biomedical ontologies is not a core skill among experimentalists. Nanopublications are not part of the mainstream publication process. Natural language processing systems cannot yet automate the process of reading research papers for us, much less derive automated databases and graphic representations of findings from these publications. Time limitations and tradition also make the prospects for collective participation in the research mapping enterprise unlikely. Disclosed herein are examples of at least three strategies for building research maps. These strategies are not mutually exclusive. The first is a publically funded data entry effort. Specialists in various fields of research could be hired to write nanopublications for papers in their field. The database of nanopublications could then be

deployed with a graphical interface. Forums, where the research community could critique the process, would be critical for the development and quality control of this effort.

[0061] The second strategy for building research maps piggybacks on activities that are part of the research community's typical workflow, such as note taking. From the time that they are students to the time that they are PI's, researchers take notes on the papers that they read. Cloud-based note taking applications (e.g. Evernote) could be used to weight, integrate and eventually share these notes. If the workflow for note taking took the form of nano-publications, papers could be transcribed into nanopublications as an automatic byproduct of researchers doing what they already do. For example, a question and answer workflow could be developed for an online pdf reader. As a user reads research articles, questions about experiments are asked, and when answered, yield a database of structured notes for the user (and everyone with access to that database). This database would be useful to the user, as a simplified record of what was read, and useful for generating research maps as well.

[0062] The third strategy for building research maps builds nano-publications into the existing publication process. Different approaches could be taken toward implementing this strategy. For example, Microsoft has developed a plugin that assists authors in using ontologies to markup their text as they write. The markup could be used to render future papers machine-readable. This would be an indirect approach. A more direct approach would incorporate fields for nano-publications into the templates for journal article submission. The NCBO makes an autocomplete widget for such purposes freely available. The widget will recommend terms from NCBO hosted ontologies when a user has started typing in a data entry form field. The nano-publications resulting from filling out these forms could be published to a public database, just as abstracts are published to PubMed. As illustrated in FIG. 1, this type of database would be the starting material for the construction of research maps.

[0063] Efforts to derive simplified representations of research findings have had neither an explicit framework nor a data infrastructure sufficient to make the approaches proposed here a cost-effective endeavor. Recent developments from neuroinformatics and machine learning can now help us to overcome these hurdles.

[0064] According to some embodiments, disclosed is a strategy to generate maps of research findings. According to some embodiments, disclosed an application that allows users to generate these maps.

[0065] According to some embodiments, two components are considered: (1) a Framework to categorize experiments, and (2) a set of algorithms to organize these experiments into “weighted network maps of experiments” (i.e., rules of Integration).

[0066] There are three basic kinds of experiments in the top hierarchy of our Framework: (1) attempts to discover new phenomena and understand their properties (**Identity Experiments**); (2) tests of causal hypotheses (**Connection Experiments**); and (3) efforts to develop and characterize new tools for performing Identity and Connection experiments (**Tool Development Experiments**).

[0067] Most Connection Experiments involve manipulating a single variable (Single Connection Experiments) and measuring the changes on another. Occasionally Connection Experiments involve simultaneously manipulating two or more variables (**Multi-Connection Experiments**). Single Connection Experiments—testing some hypothesis such as  $A$  causes  $B$  ( $A \rightarrow B$ )—come in three different sub-varieties. **Positive Manipulation Experiments** increase the probability of phenomenon  $A$  and measure for an effect on phenomenon  $B$ . For example, the use of a drug to increase the probability that a type of receptor ( $A$ ) will be active in a specific brain region, and the use of a behavioral task to measure a specific type of memory phenomenon ( $B$ ) known to be dependent on that brain region, would generally count as a Positive Manipulation experiment. The levels of the activity of  $A$  are increased, so  $A$  is positively manipulated.

[0068] **Negative Manipulations** decrease the probability of  $A$  and measure  $B$ . For example, to explore a possible causal connection between receptor  $A$  and memory phenomenon  $B$ , one could study the impact on  $B$  of a drug that blocks receptor  $A$ . It should be easy to see how the type of experiment described above (Positive Manipulation) compliments the experiment just described; the two use very different approaches to probe the hypothesized causal relation between  $A$  and  $B$ .

[0069] One problem with only using even clever combinations of Positive and Negative Manipulation experiments to test whether  $A$  affects  $B$  is that such experiments

always artificially change *A*. Therefore, any changes observed in *B* may not necessarily result from a causal connection from *A* to *B*. They could instead be experimental side effects of the artificial way *A* was changed (manipulated). The experimental process manipulating *A* may also manipulate another phenomenon (*C*) in proximity of *A*. Although *C* is *B*'s real cause, it appears to the experimenter, who is oblivious to *C*'s involvement, that *A* instead causes *B*. This possibility reveals the need for Non-Intervention experiments to supplement Positive Manipulations and Negative Manipulations.

[0070] **Non-Intervention Experiments** measure *A* and *B* without manipulating either. These experiments help us to learn whether the relationship between *A* and *B* exists outside of an experimental setting. Without these experiments, it is more difficult to be confident that our other experimental results generalize beyond the artificial manipulations. Convergent evidence among these three types of experiments (Positive and Negative Manipulations, Non-Interventions) is generally taken as good support for the hypothesis that *A* is part of the cause of *B*.

[0071] As mentioned above, the categories of connection experiments reviewed apply to **Single Connection** experiments (i.e., experiments that test one causal connection at a time). However, there are MCC experiments that manipulate multiple phenomena simultaneously and look at the effects on another phenomenon. **Multi-Connection** experiments (also called **Mediation** experiments) are simply composites of several simultaneous **Single-Connection** experiments. In general, Multi-Connection experiments help to unravel the mechanism of a single causal connection: How does *A* cause *B*? Is phenomenon *C* part of the mechanism by which *A* causes *B* ( $A \rightarrow C \rightarrow B$ )? Beyond testing the connection between *C* and *B*, to determine whether *C* is part of the mechanism by which *A* causes *B*, one would need to simultaneously manipulate *A* and *C* and then measure *B*. If *C* mediates the effects of *A* on *B* ( $A \rightarrow C \rightarrow B$ ), then Multi-Connection experiments should show that manipulations of *C* affect how changes in *A* impact *B*.

[0072] To assess the consistency and convergence of experimental results, one may attend not only to the manipulations and measurements performed but also the experimental outcomes. Therefore, experimental outcomes will be a critical component of our maps of research findings. The outcomes of Connection Experiments can only vary in a few ways, and these possibilities can be used to categorize them: A change to the probability

of  $A$  will either increase the probability of  $B$ , decrease the probability of  $B$ , or have no effect. When  $A$ 's probability or magnitude is increased experimentally and  $B$  decreases, we have evidence for an inhibitory relationship between  $A$  and  $B$ . When  $A$  is increased and  $B$  increases, evidence is provided for a facilitating relationship. When  $A$  is increased or decreased and  $B$  does not change, we have evidence for the absence of a relationship between  $A$  and  $B$ .

**[0073]** Implementation of convergence, consistency and robustness of results across different types of experiments is a principal strategy used in our maps for determining the reliability of results, and the usefulness of hypotheses. Manipulations and measurements are used to generate data. Data are analyzed to make inferences about what happened in a particular experiment and how the outcome of that experiment comes to bear on the outcomes of other experiments. These inferences yield conclusions about experimental hypotheses—for example, that  $A$  reliably affects  $B$  or that  $A$  and  $B$  are independent of each other. As stated above, the process of analyzing the result of a collection of experiments is referred to as “Integration.”

**[0074]** Methods of Integration are directed at testing the strength (i.e., the weight) of a particular causal connection (see below). These include: a) Convergency Analysis b) Consistency Analysis, which take the form of Proxy Analysis and Replication Analysis, c) Eliminative Inference, d) Mediation Analysis, and e) Robustness Analysis. Together, these different forms of Integration help us to distinguish stronger hypotheses in the experimental literature from weaker hypotheses by uncovering patterns of consistency and convergence in evidence. Therefore, these principles facilitate assembling and weighting research maps.

**[0075]** **Convergency Analysis** tests whether the outcomes of three different kinds of connection experiments (Positive Manipulations, Negative Manipulations, and Non-Interventions) are consistent with each other (i.e., whether they converge). For example, suppose we find that a drug blocks receptor  $A$  and causes a deficit in spatial learning. Suppose another drug that enhances the activity of receptor  $A$  also enhances the same form of learning. If we found that during spatial learning receptor  $A$  is activated, then our combined results would make a compelling argument that activation of receptor  $A$  is part of a cause of spatial learning. This convergence between these three types of experiments would be reflected into a higher weight for the connection in our map representing the possible relation

between *A* and *B*. On the other hand, contradictions amongst the data just outlined would weaken the weight of the connection representing the *A* to *B* relation.

[0076] According to some embodiments, a step is performed to assess whether similar experimental manipulations generally have the same kind of effect—that is, whether we have reproducibility of results (**Consistency Analysis**). For example, we might ask whether different kinds of Positive Manipulations of a given type of receptor always result in an increase on a specific behavioral measure in a specific memory task. In looking for consistency among experimental results, we can demand more or less exactness. **Proxy Analysis** determines whether different but theoretically similar Connection Experiments have the same result. In Proxy Analysis, we can abstract from the details of phenomenon *A*, phenomenon *B*, or both *A* and *B*. For example, we can ask whether genetic and pharmacological negative manipulations of receptor *A* have the same impact on spatial learning. **Replication Analysis**, on the other hand, looks for consistency among experiments that employ exactly the same variables (e.g., the same receptor agonist, applied in the same way, with results gathered using the same task measures).

[0077] Knowing that *A* affects *B* is not the same thing as knowing *how* *A* affects *B*, or what is commonly referred to as the mechanism of *A*'s effects on *B*. Understanding this mechanism (with **Mediation** experiments) increases the confidence on the causal connection between *A* and *B*. Generally, the mechanism for the  $A \rightarrow B$  causal connection involves the identification of mediators in that causal connection—that is, the go-betweens by which *A* is able to affect *B*. For example, suppose we know that spatial learning (*B*) depends on the function of some receptor (*A*). To understand how receptor *A* affects spatial learning *B*, we could start by asking what receptor *A* does that is required for spatial learning *B*. We might look for a protein (*C*) activated by receptor *A* that is also required for spatial learning *B*, such that  $A \rightarrow C \rightarrow B$ . We'll refer to this process of searching for the most likely mediators as **Mediation Analysis**, a process whose success depends heavily on Multi-Connection Experiments (*see above*). An example of a Multi-Connection experiment would be to determine whether the enhancement of memory phenomenon *B* (e.g., spatial memory) caused by a Positive Manipulation of receptor *A* is affected by a Negative Manipulation of *C*. Showing that a Negative Manipulation of *C* prevents the enhancement in *B* brought about by *A* would suggest that *C* mediates the connection between *A* and *B*. Similarly, showing that the effects of a Negative Manipulation of *A* on *B* could be prevented by a Positive Manipulation

of *C* (another example of a Multi-Connection experiment) would also be consistent with the idea that *A* causes *B* through *C*. Another method of Integration (i.e., **Eliminative Inference**) is used to assess competing hypotheses and eliminate alternative explanations for one experimental result when they conflict with the results of control experiments.

[0078] Additionally, Robustness Analysis reflects the idea that in strong causal connections a small change in *A* can have a big effect on *B*. The weaker the effect of *A* on *B* the harder it is to establish that *A* is part of the causes for *B*. Therefore, **Robustness Analysis** is an important component of determining the weight of a possible causal connection between *A* and *B*.

[0079] At the heart of most Integration methods is the familiar concept of evidential convergence—the notion that multiple, distinct lines of evidence are preferable to one line and that different types of experiments (Positive and Negative Manipulations, Non-Interventions) make unique contributions to testing the reliability of an hypothesized causal connection.

[0080] Beyond organizing experiments to be included in a research map, the framework and Integration methods introduced here can also be used to find out what additional experiments could be done to further test any connection in a causal path. One of the key practical applications of our framework will be its use for organizing experimental evidence, and thereby revealing what we know, what we do not know, what we are uncertain about, and why, at any given time. Knowing what experimental evidence we are missing is helpful in determining which experiments to perform next. The more systematic and explicit we can render this knowledge, the more thorough the basis for these choices.

[0081] At the highest zoom levels, research maps relate to a network where the nodes represent the identity and key properties of biological phenomena, and the edges represent causal connections amongst the phenomena in the map. In a research map individual phenomenon are tracked or defined in three complementary ways: *what* that phenomenon is, *where* that phenomenon exists, and *when* that phenomenon acts. In a research map, a score (see below) gives users a sense of the strength and polarity of evidence represented by each edge. These scores are calculated according to the integration rules, described above. As stated above, Convergency and consistency amongst these experiments,

for example, increases the score assigned to each edge, while contradictions in results have the opposite effect.

[0082] Additionally, symbols above the edges inform users of the types of experiments represented in each edge (Positive manipulations, Negative Manipulations, Non-interventions and Mediation experiments.). By selecting any one edge in the map, users are directed to the exact research papers and experiments represented by that edge. At the lowest zoom levels, the maps guide users to different domains in a field, while intermediate-zoom levels reveal topics represented in the map within each domain.

[0083] According to some embodiments, to assemble a research map, trained users initially extract from published research papers findings that describe the identity of biological phenomena (the nodes in the map), as well as those experiments that test causal connections between these phenomena (the edges in the maps). Using manually entered examples, machine-learning routines will then systematically populate the maps with similar and related experiments by crawling multiple resources such as manuscripts associated with the Library of Medicine. With such a map, neuroscientists could instantly evaluate the amount and type of evidence available for any one causal connection in the map. Research maps can be structured and machine readable, and users will be able to interact with the maps dynamically. For example, they will be able to query them for possible connections between any two phenomena of interest, mine them for hitherto unsuspected relations and for micro and macro trends. Moreover users can generate personalized private maps with their own unpublished results.

[0084] According to some embodiments, information pertaining to a research paper is stored in a graph. Each node in the map represents an *Agent* or a *Target*. An *Agent* is a phenomenon that is changed or observed in an experiment and that acts, or putatively or potentially acts, on another phenomenon (the *Target*), or whose action, if any, on the *Target* is to be determined or potentially determined in the experiment. A *Target* is a phenomenon whose change we measure is caused, or potentially caused, as a result of a change in the *Agent*. As described above, each *Agent* and *Target* are described by three properties: a) *What* describes a key identifier of the phenomenon involved (e.g., the name for which the gene, protein, cell, organ, behavior is known); b) *Where* describes the location of the *What* (e.g., the cellular, organ, species where the *What* in question); and c) *When* - Pertains to temporal



information that is critical for the identity of the *What* (e.g., the time/age/phase). For example, we may measure the protein neurofibromin (*What*) in different locations, which would result in different identities for neurofibromin since this protein could have different biological properties in different locations or at different times in development.

[0085] The four different types of experiments are represented by different symbols above the edges representing connections amongst phenomena in the map: **Positive Manipulations** are represented by an up arrow ( $\uparrow$ ), **Negative Manipulations** are represented by a down arrow ( $\downarrow$ ), **Non-Interventions** are represented by a crossed zero ( $\phi$ ), and **Mediation Experiments** are represented by a downward triangle ( $\nabla$ ) (figure 1).

[0086] From the 4 different experiments described above, we can glean 3 different types of relationships between an *Agent* and a *Target*: Connections amongst phenomena are defined as *Excitatory* when increases in the *Agent* lead to increases in the *Target* while decreases in the *Agent* lead to decreases in the *Target*. In an excitatory relation a Positive Manipulation experiment would result in an increase in the *Target* and a Negative Manipulation experiment would result in a decrease of the *Target*.

[0087] In Inhibitory connections a Positive Manipulation experiment would result in a decrease in the *Target* (hence the name) and a Negative Manipulation experiment would result in an increase of the *Target*. When manipulations of the *Agent* fail to affect the *Target*, this is evidence for the absence of a connection between the two phenomena. Non-connections can also be represented in a research map (see FIG. 6).

[0088] According to some embodiments, as shown in FIG. 6, each node 610 in the graph 600 can have items 630 that describe, for example, the name of the node 610 (what; top), as well as spatial (where; middle) and temporal (when; bottom) information that defines it. Nodes can include agents 612 and targets 614. Nodes 610 are connected by edges 620 that define the nature of the causal relations represented, including excitatory (sharp edges 620a), inhibitory (dull edge 620b) and no relation (dotted line 620c). Each edge 620 also has a score 640 that reflects the amount of evidence represented, and symbols 650 that reflect the types of experiments carried out (e.g., upward arrow for Positive Manipulations, downward arrow for Negative Manipulations, and triangle for Mediation Experiments). As will be

appreciated, any of a variety of symbols may be used consistently. A legend may be provided to correlate a symbol with its meaning.

[0089] According to some embodiments, to assign a score 640 to each of the edges 620 in a research map, the Integration principles described above can be used. These principles reflect epistemological rules commonly used in many fields in biology, including molecular and cellular biology, cancer, immunology, neuroscience, etc. Convergent and/or consistent results increase the score of an edge representing an experimental assertion, while conflicting results decrease that score. Each experiment category (Positive Manipulation, Negative Manipulation, Non-Intervention and Mediation) contributes a maximum of 0.25 to the overall score. Other maximums can be used where weighted consideration of categories is desired. Multiple experiments of the same kind contribute increasingly smaller scores to the edge. To calculate the rate at which additional experiments of the same type (e.g., a Positive manipulation experiment) contribute to the score, a geometric progression is used with the start term = 0.125 and  $r = 0.5$ . The example in Table 1 describes an excitatory relation between two nodes in a map. It represents three experiments (Exp 1, Exp 2 and Exp 3 in Table 1) for each of the four types of experiments described above, to yield a set of 12 scores. The overall score for the group of experiments represented in Table 1 (assuming there are no conflicting results) is 0.875. This would be taken as very compelling evidence that there is an excitatory relation between the referenced Agent 612 and Target 614.

Type	Exp 1	Exp 2	Exp 3
Positive	0.125	0.0625	0.03125
Negative	0.125	0.0625	0.03125
Mediation	0.125	0.0625	0.03125
Non-Intervention	0.125	0.0625	0.03125

**Table 1: Overall Score**

[0090] According to some embodiments, contradicting evidence can weaken a hypothesis. Experiments that contradict the results of other experiments decrease the score

assigned to the edge(s) representing those experiments, such as for an example shown in Table 2.

Edge Type	Exp 1	Exp 2	Total Score
Excitatory	0.125	0.0625	0.1875
Inhibitory	0.125		0.125
No Connection	0.125		0.125

**Table 2: Resolving Contradictory Scores with a Dominant Connection Type**

**[0091]** From the results summarized in Table 2, it can be understood that there is more evidence for an excitatory connection (the score is higher than the other three scores for the alternative connection types). However, one may not assign a score of 0.1875 to that excitatory connection because there is contradictory evidence represented by the other two types of connections listed in Table 2 (inhibitory and no-connection).

**[0092]** According to some embodiments, to account for contradictory results, operations illustrated in FIG. 7 may be performed. For example, in operation 710, the connection type is determined with the highest calculated score (the “max” score); in the case outlined in Table 2 this would be the excitatory connection with a score of 0.1875. In operation 720, the “overall score” is calculated as the sum of all scores for the connections representing that specific edge. In the example above, this would be 0.4375. In operation 730, these two values allow recalibration of the score for the dominant connection taking into account the contradictory evidence outlined in Table 2. This can be done with the following formula:

$$\text{Connection score} = \frac{\text{Max}^2}{\text{Overall}}$$

Hence, for the case in Table 2, the score for the excitatory connection (the dominant connection) would be 0.0803.

[0093] For the exemplary scenario described in Table 3, none of the hypotheses concerning the type of connection being studied is supported by more evidence than the others: each of the three hypotheses is supported by equal but contradictory evidence. Therefore, in this unique case, there is no predominant evidence for a given edge type (excitatory, inhibitory, or no connection), and consequently, the system does not represent an edge between that Agent and Target pair.

Edge Type	Exp 1	Total Score
Excitatory	0.125	0.125
Inhibitory	0.125	0.125
No Connection	0.125	0.125

**Table 3: Resolving Contradictory Scores Without a Dominant Connection Type**

[0094] According to some embodiments, a web application implementing research maps can be hosted by Amazon Web Services (AWS), Elastic Compute Cloud (EC2) platform, with an **Ubuntu 12.04 64-bit** operating system. For the backend, **node.js** can be used for its single-threaded event loop and callback-based model. For the database, **neo4j**, a NoSQL graph database, can be used. Graph databases store data in nodes and edges compared to tuples in relational databases. With graph databases, performance of recursive queries is not a bottleneck. To query the database, the **Cypher Query Language (CQL)** of neo4j can be used. To display the graphs, **Graphviz** can be used. **PubMed's** interface (<http://eutils.ncbi.nlm.nih.gov>) can be used to retrieve information about papers and **NIF** (Neuroscience Information Framework, <http://nif-services.neuinfo.org>) to generate correct ontologies for the user.

[0095] According to some embodiments, as shown in FIG. 8, experiments can be added to the app by way of an interface 800 for building a research map. FIG. 8 shows the interface of the app used to enter experiments by inputs 810. The top of the figure shows the citation 820 for the research article that served as the source of the experiments in the research map 830 shown on the right. The left panel of the figure shows the interface with inputs 810 for entering the details of the experiments.

[0096] FIG. 8 illustrates the interface used to enter the details of experiments, including the identity of the *Agent* and *Target* including the “What”, “Where” and “When” fields for items 630 (see above), the type of “Manipulation”, the type of “Result”, and succinct descriptions of the “Approaches” used to manipulate the *Agent* and measure changes in the *Target*. Upon hitting “Submit”, the app returns a map 830 similar to the one shown on the right panel of FIG. 8.

[0097] According to some embodiments, in the internal database representation of experiments entered into the app, a User node is connected to a Paper node, which is connected to the Experiment node(s) in that paper. Each Experiment node can be connected to two NeuroLaxTerm nodes representing the *Agent* and the *Target* for that particular connection (edge). *Agent* and *Target* can be connected in a neo4j representation.

[0098] According to some embodiments, natural language processing and machine learning algorithms can be used to automate the process of entering experiments into the database of research maps. By documenting the provenance of each entry in the map (*what, where, when, manipulation approach results*, etc.) with the appropriate text in the original research article, the app will be able to automatically enter other experiments similar to those that were manually entered. The text highlighted by the user will be used to find other similar experiments in research articles in sources such as the Library of Medicine. Users can then check the accuracy of the experiments automatically entered by the app, and thus further improve the future accuracy of these processes: machine learning algorithms get better with increased usage and feedback.

[0099] According to some embodiments, as shown in FIG. 9, the research maps app also has an interface 900 where users can search the database for specific terms, such as CREB as an example in FIG. 9. The search can process information supplied to inputs 910 to generate a research map 930. The search is made using the NeuroLaxTerm nodes discussed above (when available). The user can specify parameters like minimum and maximum path scores (the weights of each edge), as well as the number of hops (i.e., how many consecutive edges) the search will take from the search term (in this case CREB). neo4j allows for more efficient searches than mySQL since the same functionality in mySQL requires recursive queries which need to go through the entire database for each recursion. In contrast, in neo4j

the same problem becomes a graph traversal. Using Breadth First Search, the system need not reference the entire database, only through the degree of the vertex in question.

**[0100]** While interacting with the inputs 910, users can introduce constraints on the resulting graphs of the research map 930, such as minimum and maximum path scores. This and other parameters, allow users to constrain the connections visualized in the resulting graphs. By limiting the graphs to edges with high scores, for example, users can visualize only those connections with the highest levels of evidence (e.g., those that are likely to be more reliable). Similarly, by limiting the graphs to edges with low scores users can quickly identify those experimental assertions with the least amount of evidence (i.e., with the greatest need of further investigation).

**[0101]** According to some embodiments, users can also combine different nodes in a graph, and the app will redraw the graph accordingly. This is useful for combining data from related nodes (*Agents or Targets*) that were entered as separate entities. This and other functionality in the app can be applied as a hypothesis building tool, allowing users to explore the causal ramifications of different hypotheses or assumptions.

**[0102]** According to some embodiments, users can determine the provenance of any one edge on these combined maps. Selecting any edge can direct the user to a table (see FIG. 10) with all of the experiments represented in that edge. The resulting table gives users the option of being redirected to the experiments and research papers represented in that edge.

**[0103]** According to some embodiments, as shown in FIG. 9, an interface 900 of the app is used to interact with the data in the app. The left side of FIG. 9 shows the app panel used for entering inputs 910 of the details for a particular query (in this case an exemplary query is directed at the protein CREB). The research map 930 shown on the right is only a fraction of the map that this query returned. This map represents data from many different research articles

**[0104]** According to some embodiments, as shown in FIG. 10, a user can connect to each experiment in an edge of a research map 1030 via an interface 1000. The top of FIG. 10 shows the mode selection 1020 of the app used to interact with the data in the app. Based on the mode selection 1020, a search can be performed with respect to agents, targets, or

both. The left side of FIG. 10 shows the app panel of inputs 101 used for entering the details for a particular query (in this case the query was directed at the protein CREB). The research map 1030 shown on the right is only a fraction of the map that this query returned.

[0105] By defining both the Agent and the Target, users can explore causal paths between any two phenomena in the database of the app. This and other aspects of the search functionality of the app aid users in both integrating previous findings and planning future experiments. By simply defining a couple of terms and parameters users can get a snap shot of a lot of complex results potentially coming from a considerable number of research papers. At a glance, users can glean from research maps not only the type of connection between any given Agent and its Target, but also the amount and type of evidence that supports that connection. By representing with clear symbols the kinds of experiments on each edge on the map, as well as the corresponding scores, users can quickly get critical information about a particular area of research. Review articles are the traditional media for summarizing specific bodies of research. However, they are not concise, objective or dynamic, they cannot easily be updated and they are not interactive, all features and strengths of research maps. What a user may be able to learn in minutes with research maps, could otherwise take days to glean from traditional research article searches.

[0106] According to some embodiments, an interactive website can be created to receive inputs from a user and provide outputs. The website can be backed by a relational database, with user authentication and authorization of content for submission, viewing, and editing. According to some embodiments, a mechanism of data entry is provided for experiments according to S2 Framework as discussed in the book Engineering the Next Revolution in Neuroscience, association of entered experiments with published or unpublished scientific articles, and graphical visualization and filtering of the network of biological phenomena described by the experiments. The application presents the user with multiple views of the data, both graphical and tabular, to help the user navigate the network of experiments (NEX). The application uses an algorithm to generate scores that summarize the significance of recorded experiments connecting pairs of phenomena and make the graphs simpler to understand.

[0107] According to some embodiments, the website enables visitors to register an account by providing an email address and password. Registered users can login to the

website by providing a valid email address and password. The website provides users a My Profile page with content they can edit including textual fields for name, affiliation, website, and about section. The website provides each user the option to be either visible to other users or to be invisible to them, and to change this setting whilst editing the My Profile page.

[0108] According to some embodiments, the website provides each user with a My Article Entries page that lists that user's Article Entries with thumbnail images of corresponding graphs. Clicking on an item takes the user to the page for that Article Entry. Users can view their own My Article Entries pages as well as those of other users if the other users have chosen to be visible. The My Article Entries page for the logged-in user also provides a search bar that supplies matching article results from PubMed in a drop-down menu for users to click on a selection to begin an Article Entry about that article, as well as a form input field for the logged-in user to begin an Article Entry about an article that is not on PubMed by supplying a title for the new Article Entry into the input field and submitting the form.

[0109] According to some embodiments, the website provides a separate page for each Article Entry. An Article Entry page is visible to the authoring user as well as users other than the authoring user if the authoring user is visible and the authoring user has clicked "Publish" for that Article Entry.

[0110] According to some embodiments, the Article Entry page is divided into areas for content of different purposes. The header shows the PubMed article title and citation information, or the title submitted by the user if it is not a PubMed article. When viewed by the authoring user, the header also includes a link to edit the title if it is not a PubMed article, and buttons to Publish or Un-publish or Delete the Article Entry. If the Article Entry is not published, a left column below the header contains a form to enter Experiments about the article.

[0111] According to some embodiments, the Experiment form contains the following fields: Agent, Manipulation, Approach Used, Target, Measurement, Approach Used, Brain Location of the Studies, Developmental Stage of the Studies, and Other Info. As well, a user can elect to enter additional Agents, as many as needed to describe Multi-Connection Experiments. The required fields are Agent, Manipulation, Target, and Measurement. Permitted values of the Manipulation and Measurement fields are constrained



to 'Positive', 'Negative', or 'None'. The textual fields feature suggestions that populate in a drop-down menu as the user types. The Agent and Target suggestions are aggregated from matching Agent and Target terms that a user has previously entered, and matching terms from Neuroscience Information Framework (NIF). The Approach Used fields feature a similar suggestions mechanism, except that the suggestions are specific to previously entered values for Approach Used plus suggestions from NIF. The Brain Location and Developmental Stage fields are likewise. When suggestions from NIF are selected, the website stores the resource information to use for matching terms having identical resources to build the graphs. When a user does not select a suggestion and instead enters a term with no resource, the website matches it to other terms entered by that user by the term's name only. Once entered, experiments for an Article Entry can be duplicated, edited, and deleted by the authoring user via like-named links within the Article Entry page.

**[0112]** According to some embodiments, the website can provide for side-by-side experiment entry while viewing the article abstract and meta-data including keywords. The layout facilitates the data-entry process that can involve copy-paste operations by putting much of what a user needs on a single page, and respecting copyright and license for the full-text article content despite when programmatically available by re-publishers such PubMed Central. The article references are rendered as links such that a user can begin a new Article Entry for each article reference by clicking on the reference.

**[0113]** According to some embodiments, another display shows multiple graphical and tabular views of the experiments that have been entered. There are three basic views of the experiments. The first, called "Integrated", shows the directed graph of the network of phenomena wherein edges represent the presence of reported experiments that concern the connected phenomena meeting certain criteria/rules. The Integrated graph also shows scores for each edge representing the strength of that experimental connection as determined by an algorithm. The scores are used to color the edges according to a red-yellow-blue heat map. The second view of experiments, called "Raw", shows a separate directed graph for each pair of phenomena that are connected by reported experiments, with each edge representing exactly one reported experiment. A table is provided that summarizes the information entered about the experiments and that allows the user to highlight edges in the graph by hovering over rows in the table using the mouse. The third view depicts single experiment diagrams/maps as used in the book "Engineering the Next Revolution in

Neuroscience". Users navigate the three views by clicking on edges in the graphs or clicking on textual navigation links, or by scrolling. As such, all of the NEX content can be linked to via the URL.

[0114] According to some embodiments, the rules for display of edges in the Integrated graph are that edges are only shown when the experimental relationship determined by the scoring algorithm is either Facilitating or Inhibitory, i.e., Non-Causal edges are not shown nor those of score value zero. This rule is designed to simplify the Integrated view for the user.

[0115] According to some embodiments, the scores can be designed to have values between 0 and 1. To perform Consistency Analysis and Convergency Analysis according to method disclosed herein, the algorithm counts numbers of repetitions of single-connection experiments having Positive Manipulations, Negative Manipulations, and Non-Interventions for each pair of connected phenomena. The website need not make a distinction between Consistency by direct Repetition versus Consistency by Proxy. The application can identify such differences eventually by utilizing the Approach Used fields to match experiments with identical or different sets of values. Experiments of the three manipulation types are scored using geometrically decreasing sums such that an infinite number of repetitions of Positive Manipulation experiments between one Agent and Target phenomenon pair would generate a score of 0.25 for that edge (i.e. the first such manipulation would have score 0.125 and each successive experiment would count half as much as the preceding one). Thus, the score contribution of single-connection experiments could be at most 0.75 after accounting for the three kinds of manipulations (Convergency Analysis) and their repetitions (Consistency Analysis).

[0116] Since different experimental results support different kinds of experimental relationships, a score can be computed for experiments that support each of the relationships Facilitating, Inhibiting, and Non-Causal. The dominant relationship is determined by maximum score and from its score is subtracted that of the next sub-dominant score. Therefore, when the experimental record as integrated by Convergency and Consistency Analysis equally supports two contradictory relationships, the edge will have a score of zero. When a dominant relationship is present, the last up to 0.25 of its score is calculated by a diminishing geometric sum of mediation experiments (Mediation Analysis).

Mediation experiments are multi-connection experiments, those having more than one agent, for which the experiment reveals the additional agents to have a mediating effect, i.e. it interferes with the relationship determined by single-connection experiments. It is contemplated that the scoring algorithm will evolve in particular to Integrate other methods of Analysis such as Robustness and Eliminative Inference as described in Engineering the Next Revolution in Neuroscience.

[0117] According to some embodiments, the website provides a My Map interface that is similar in appearance to the Article Entry page. This page features filters in place of the Experiment entry form. The filters allow the user to render NEX visualizations composed of multiple Article Entries. The filters allow the user to construct and navigate collections of experiments according that meet criteria. Filters include My Article Entries, Shared Article Entries, Date Article Entry Published, Article Entry Author, Article Authors, Date Article Published, Article Journal, Phenomena of Interest, Brain Locations, Developmental Stages, Manipulation Type, Presence of Contradiction, and Score Value.

[0118] According to some embodiments, the website is built using a model-view-controller framework to process web requests. The website can store data in a relational database with the attached schema. For example, the website may be built atop the Ruby on Rails framework and MySQL database which are industry standards. The website generates queries to PubMed.gov using PubMed Entrez Programming Utilities to supply users with article search results and to obtain meta-data about articles including article citations. PubMed is the most comprehensive and government-curated database of research article citations. The website generates JavaScript queries to Neuroscience Information Framework (NIF, neuinfo.org) using its OntoQuest Web Services to supply suggestions for scientific terms as users enter experiments. The NIF ontologies currently provide the most comprehensive resource for neuroscience terms. The website generates graphical images of networks using command-line tools of the Graphviz open source graph visualization software. The website uses the Scalable Vector Graphics output format of Graphviz and embeds the SVG graphics inline with the HTML. The specific technologies used for graphical visualization of the networks are expected to change as more Users join, Article Entries and Experiments are contributed, and the maps grow in size.

[0119] The subject technology can be applied to any of a variety of fields. Beyond helping biologists integrate and plan experiments, research maps of the subject technology can also have a key role in the process of reviewing and publishing science. Despite the best efforts of all involved, the process of reviewing and publishing science is fraught with subjectivity and arbitrary judgments that can compromise the fairness of the reviews and the ultimate goal of rewarding the best and most promising science. Research maps of the subject technology provide a means to not only help objectively gaging the quality of any body of research work, but also to help guide decisions about the potential promise of proposed experiments. The clarity and perspective afforded by visual graphical summaries of complex bodies of research data can help reviewers in the complex and often agonizing process of reviewing research products and proposals. Ultimately, research maps can be defined by a geometrical complexity that captures the causal structure of a research paper, area, etc. This geometrical complexity could be defined numerically, and thus it will be possible to also gauge objectively the contribution that a specific body of work has had or could have on any one given map. For example, an important/significant body of work should have a substantial impact on the geometry and edge weights of a given research map, while a less impactful set of experiments would have a comparatively modest effect on these same measures. Research contributions could have innovation scores associated with them that reflect their impact on the geometry of research maps. Like all measures used in evaluating science, the measures based on impact on research maps would have to be used in the context of other considerations. Convergence and consistency of evidence can be evaluated. In this respect, research maps of the subject technology would be a worthwhile addition to the rocky and temperamental process of reviewing and publishing scientific articles.

[0120] The principles used in the subject technology can be applied to any of a variety of fields. The principles that govern the detection and structure of causation in science, such as Convergency, Consistency and Robustness, can be generally useful. In an age of proliferation of information of different quality and dependability, the approaches used in embodiments of the subject technology could be used to ferret out the potential causal structure of entities that shape our political and economic worlds. Research maps of the subject technology could be used together with other strategies to advance the noble goals of the Semantic Web. Information, whether in science, politics or economics, is likely to play

by the same rules. Therefore, the simple principles behind research maps of the subject technology could be used to bring structure and order to an abundance of information.

[0121] A phrase such as “an aspect” does not imply that such aspect is essential to the subject technology or that such aspect applies to all configurations of the subject technology. A disclosure relating to an aspect may apply to all configurations, or one or more configurations. An aspect may provide one or more examples of the disclosure. A phrase such as “an aspect” may refer to one or more aspects and vice versa. A phrase such as “an embodiment” does not imply that such embodiment is essential to the subject technology or that such embodiment applies to all configurations of the subject technology. A disclosure relating to an embodiment may apply to all embodiments, or one or more embodiments. An embodiment may provide one or more examples of the disclosure. A phrase such as “an embodiment” may refer to one or more embodiments and vice versa. A phrase such as “a configuration” does not imply that such configuration is essential to the subject technology or that such configuration applies to all configurations of the subject technology. A disclosure relating to a configuration may apply to all configurations, or one or more configurations. A configuration may provide one or more examples of the disclosure. A phrase such as “a configuration” may refer to one or more configurations and vice versa.

[0122] FIG. 12 is a conceptual block diagram illustrating an example of a system, in accordance with various aspects of the subject technology. A system 1201 may be, for example, a client device or a server. The system 1201 may include a processing system 1202. The processing system 1202 is capable of communication with a receiver 1206 and a transmitter 1209 through a bus 1204 or other structures or devices. It should be understood that communication means other than busses can be utilized with the disclosed configurations. The processing system 1202 can generate audio, video, multimedia, and/or other types of data to be provided to the transmitter 1209 for communication. In addition, audio, video, multimedia, and/or other types of data can be received at the receiver 1206, and processed by the processing system 1202. Components of the system 1201 may include, as shown in FIG. 13, an input module 1302, an output module 1304, a map generating module 1306, a hypothesis module 1308, and a database module 1310. Each module may be operated by one or more processors. The modules may be interconnected and communicate with each other.

[0123] The processing system 1202 may include a processor for executing instructions and may further include a machine-readable medium 1219, such as a volatile or non-volatile memory, for storing data and/or instructions for software programs. The instructions, which may be stored in a machine-readable medium 1210 and/or 1219, may be executed by the processing system 1202 to control and manage access to the various networks, as well as provide other communication and processing functions. The instructions may also include instructions executed by the processing system 1202 for various user interface devices, such as a display 1212 and a keypad 1214. The processing system 1202 may include an input port 1222 and an output port 1224. Each of the input port 1222 and the output port 1224 may include one or more ports. The input port 1222 and the output port 1224 may be the same port (e.g., a bi-directional port) or may be different ports.

[0124] The processing system 1202 may be implemented using software, hardware, or a combination of both. By way of example, the processing system 1202 may be implemented with one or more processors. A processor may be a general-purpose microprocessor, a microcontroller, a Digital Signal Processor (DSP), an Application Specific Integrated Circuit (ASIC), a Field Programmable Gate Array (FPGA), a Programmable Logic Device (PLD), a controller, a state machine, gated logic, discrete hardware components, or any other suitable device that can perform calculations or other manipulations of information.

[0125] A machine-readable medium can be one or more machine-readable media. Software shall be construed broadly to mean instructions, data, or any combination thereof, whether referred to as software, firmware, middleware, microcode, hardware description language, or otherwise. Instructions may include code (e.g., in source code format, binary code format, executable code format, or any other suitable format of code).

[0126] Machine-readable media (e.g., 1219) may include storage integrated into a processing system, such as might be the case with an ASIC. Machine-readable media (e.g., 1210) may also include storage external to a processing system, such as a Random Access Memory (RAM), a flash memory, a Read Only Memory (ROM), a Programmable Read-Only Memory (PROM), an Erasable PROM (EPROM), registers, a hard disk, a removable disk, a CD-ROM, a DVD, or any other suitable storage device. Those skilled in the art will recognize how best to implement the described functionality for the processing system 1202. According to one aspect of the disclosure, a machine-readable medium is a computer-

readable medium encoded or stored with instructions and is a computing element, which defines structural and functional interrelationships between the instructions and the rest of the system, which permit the instructions' functionality to be realized. In one aspect, a machine-readable medium is a non-transitory machine-readable medium, a machine-readable storage medium, or a non-transitory machine-readable storage medium. In one aspect, a computer-readable medium is a non-transitory computer-readable medium, a computer-readable storage medium, or a non-transitory computer-readable storage medium. Instructions may be executable, for example, by a client device or server or by a processing system of a client device or server. Instructions can be, for example, a computer program including code.

[0127] An interface 1216 may be any type of interface and may reside between any of the components shown in FIG. 12. An interface 1216 may also be, for example, an interface to the outside world (e.g., an Internet network interface). A transceiver block 1207 may represent one or more transceivers, and each transceiver may include a receiver 1206 and a transmitter 1209. A functionality implemented in a processing system 1202 may be implemented in a portion of a receiver 1206, a portion of a transmitter 1209, a portion of a machine-readable medium 1210, a portion of a display 1212, a portion of a keypad 1214, or a portion of an interface 1216, and vice versa.

[0128] FIG. 11 illustrates a simplified diagram of a system 1100, in accordance with various embodiments of the subject technology. The system 1100 may include one or more remote client devices 1102 (e.g., client devices 1102a, 1102b, 1102c, and 1102d) in communication with a server computing device 1106 (server) via a network 1104. In some embodiments, the server 1106 is configured to run applications that may be accessed and controlled at the client devices 1102. For example, a user at a client device 1102 may use a web browser to access and control an application running on the server 1106 over the network 1104. In some embodiments, the server 1106 is configured to allow remote sessions (e.g., remote desktop sessions) wherein users can access applications and files on the server 1106 by logging onto the server 1106 from a client device 1102. Such a connection may be established using any of several well-known techniques such as the Remote Desktop Protocol (RDP) on a Windows-based server.

[0129] By way of illustration and not limitation, in one aspect of the disclosure, stated from a perspective of a server side (treating a server as a local device and treating a

client device as a remote device), a server application is executed (or runs) at a server 1106. While a remote client device 1102 may receive and display a view of the server application on a display local to the remote client device 1102, the remote client device 1102 does not execute (or run) the server application at the remote client device 1102. Stated in another way from a perspective of the client side (treating a server as remote device and treating a client device as a local device), a remote application is executed (or runs) at a remote server 1106.

**[0130]** By way of illustration and not limitation, a client device 1102 can represent a computer, a mobile phone, a laptop computer, a thin client device, a personal digital assistant (PDA), a portable computing device, or a suitable device with a processor. In one example, a client device 1102 is a smartphone (e.g., iPhone, Android phone, Blackberry, etc.). In certain configurations, a client device 1102 can represent an audio player, a game console, a camera, a camcorder, an audio device, a video device, a multimedia device, or a device capable of supporting a connection to a remote server. In one example, a client device 1102 can be mobile. In another example, a client device 1102 can be stationary. According to one aspect of the disclosure, a client device 1102 may be a device having at least a processor and memory, where the total amount of memory of the client device 1102 could be less than the total amount of memory in a server 1106. In one example, a client device 1102 does not have a hard disk. In one aspect, a client device 1102 has a display smaller than a display supported by a server 1106. In one aspect, a client device may include one or more client devices.

**[0131]** In some embodiments, a server 1106 may represent a computer, a laptop computer, a computing device, a virtual machine (e.g., VMware® Virtual Machine), a desktop session (e.g., Microsoft Terminal Server), a published application (e.g., Microsoft Terminal Server) or a suitable device with a processor. In some embodiments, a server 1106 can be stationary. In some embodiments, a server 1106 can be mobile. In certain configurations, a server 1106 may be any device that can represent a client device. In some embodiments, a server 1106 may include one or more servers.

**[0132]** In one example, a first device is remote to a second device when the first device is not directly connected to the second device. In one example, a first remote device



may be connected to a second device over a communication network such as a Local Area Network (LAN), a Wide Area Network (WAN), and/or other network.

[0133] When a client device 1102 and a server 1106 are remote with respect to each other, a client device 1102 may connect to a server 1106 over a network 1104, for example, via a modem connection, a LAN connection including the Ethernet or a broadband WAN connection including DSL, Cable, T1, T3, Fiber Optics, Wi-Fi, or a mobile network connection including GSM, GPRS, 3G, WiMax or other network connection. A network 1104 can be a LAN network, a WAN network, a wireless network, the Internet, an intranet or other network. A network 1104 may include one or more routers for routing data between client devices and/or servers. A remote device (e.g., client device, server) on a network may be addressed by a corresponding network address, such as, but not limited to, an Internet protocol (IP) address, an Internet name, a Windows Internet name service (WINS) name, a domain name or other system name. These illustrate some examples as to how one device may be remote to another device. But the subject technology is not limited to these examples.

[0134] According to certain embodiments of the subject technology, the terms “server” and “remote server” are generally used synonymously in relation to a client device, and the word “remote” may indicate that a server is in communication with other device(s), for example, over a network connection(s).

[0135] According to certain embodiments of the subject technology, the terms “client device” and “remote client device” are generally used synonymously in relation to a server, and the word “remote” may indicate that a client device is in communication with a server(s), for example, over a network connection(s).

[0136] In some embodiments, a “client device” may be sometimes referred to as a client or vice versa. Similarly, a “server” may be sometimes referred to as a server device or vice versa.

[0137] In some embodiments, the terms “local” and “remote” are relative terms, and a client device may be referred to as a local client device or a remote client device, depending on whether a client device is described from a client side or from a server side, respectively. Similarly, a server may be referred to as a local server or a remote server,

depending on whether a server is described from a server side or from a client side, respectively. Furthermore, an application running on a server may be referred to as a local application, if described from a server side, and may be referred to as a remote application, if described from a client side.

[0138] In some embodiments, devices placed on a client side (e.g., devices connected directly to a client device(s) or to one another using wires or wirelessly) may be referred to as local devices with respect to a client device and remote devices with respect to a server. Similarly, devices placed on a server side (e.g., devices connected directly to a server(s) or to one another using wires or wirelessly) may be referred to as local devices with respect to a server and remote devices with respect to a client device.

[0139] As used herein, the word “module” refers to logic embodied in hardware or firmware, or to a collection of software instructions, possibly having entry and exit points, written in a programming language, such as, for example C++. A software module may be compiled and linked into an executable program, installed in a dynamic link library, or may be written in an interpretive language such as BASIC. It will be appreciated that software modules may be callable from other modules or from themselves, and/or may be invoked in response to detected events or interrupts. Software instructions may be embedded in firmware, such as an EPROM or EEPROM. It will be further appreciated that hardware modules may be comprised of connected logic units, such as gates and flip-flops, and/or may be comprised of programmable units, such as programmable gate arrays or processors. The modules described herein are preferably implemented as software modules, but may be represented in hardware or firmware.

[0140] It is contemplated that the modules may be integrated into a fewer number of modules. One module may also be separated into multiple modules. The described modules may be implemented as hardware, software, firmware or any combination thereof. Additionally, the described modules may reside at different locations connected through a wired or wireless network, or the Internet.

[0141] In general, it will be appreciated that the processors can include, by way of example, computers, program logic, or other substrate configurations representing data and instructions, which operate as described herein. In other embodiments, the processors can include controller circuitry, processor circuitry, processors, general purpose single-chip or

multi-chip microprocessors, digital signal processors, embedded microprocessors, microcontrollers and the like.

[0142] Furthermore, it will be appreciated that in one embodiment, the program logic may advantageously be implemented as one or more components. The components may advantageously be configured to execute on one or more processors. The components include, but are not limited to, software or hardware components, modules such as software modules, object-oriented software components, class components and task components, processes methods, functions, attributes, procedures, subroutines, segments of program code, drivers, firmware, microcode, circuitry, data, databases, data structures, tables, arrays, and variables.

[0143] The foregoing description is provided to enable a person skilled in the art to practice the various configurations described herein. While the subject technology has been particularly described with reference to the various figures and configurations, it should be understood that these are for illustration purposes only and should not be taken as limiting the scope of the subject technology.

[0144] There may be many other ways to implement the subject technology. Various functions and elements described herein may be partitioned differently from those shown without departing from the scope of the subject technology. Various modifications to these configurations will be readily apparent to those skilled in the art, and generic principles defined herein may be applied to other configurations. Thus, many changes and modifications may be made to the subject technology, by one having ordinary skill in the art, without departing from the scope of the subject technology.

[0145] It is understood that the specific order or hierarchy of steps in the processes disclosed is an illustration of exemplary approaches. Based upon design preferences, it is understood that the specific order or hierarchy of steps in the processes may be rearranged. Some of the steps may be performed simultaneously. The accompanying method claims present elements of the various steps in a sample order, and are not meant to be limited to the specific order or hierarchy presented.

[0146] As used herein, the phrase “at least one of” preceding a series of items, with the terms “and” or “or” to separate any of the items, modifies the list as a whole, rather

than each member of the list (i.e., each item). The phrase “at least one of” does not require selection of at least one item; rather, the phrase allows a meaning that includes at least one of any one of the items, and/or at least one of any combination of the items, and/or at least one of each of the items. By way of example, the phrases “at least one of A, B, and C” or “at least one of A, B, or C” each refer to only A, only B, or only C; any combination of A, B, and C; and/or at least one of each of A, B, and C.

[0147] Furthermore, to the extent that the term “include,” “have,” or the like is used in the description or the claims, such term is intended to be inclusive in a manner similar to the term “comprise” as “comprise” is interpreted when employed as a transitional word in a claim.

[0148] The word “exemplary” is used herein to mean “serving as an example, instance, or illustration.” Any embodiment described herein as “exemplary” is not necessarily to be construed as preferred or advantageous over other embodiments.

[0149] A reference to an element in the singular is not intended to mean “one and only one” unless specifically stated, but rather “one or more.” Pronouns in the masculine (e.g., his) include the feminine and neuter gender (e.g., her and its) and vice versa. The term “some” refers to one or more. Underlined and/or italicized headings and subheadings are used for convenience only, do not limit the subject technology, and are not referred to in connection with the interpretation of the description of the subject technology. All structural and functional equivalents to the elements of the various configurations described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and intended to be encompassed by the subject technology. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the above description.

[0150] While certain aspects and embodiments of the invention have been described, these have been presented by way of example only, and are not intended to limit the scope of the invention. Indeed, the novel methods and systems described herein may be embodied in a variety of other forms without departing from the spirit thereof. The accompanying claims and their equivalents are intended to cover such forms or modifications as would fall within the scope and spirit of the invention.

## WHAT IS CLAIMED IS:

1. A method comprising:
  - receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment;
  - by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target;
  - wherein the research map comprises nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges;
  - wherein one of the nodes represents the selected agent or the selected target;
  - wherein each of the edges comprises an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and
  - outputting, by a processor, a hypothesis positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.
2. The method of claim 1, further comprising generating a hypothesis map based on the hypothesis.
3. The method of claim 2, further comprising displaying a graphical representation of the hypothesis map on a display device.
4. The method of claim 1, further comprising displaying a graphical representation of the research map on a display device.
5. The method of claim 1, wherein the indication of the causal relationship comprises a strength of the causal relationship.
6. The method of claim 1, wherein the indication of the causal relationship comprises a positive or negative correlation between the connected nodes.
7. The method of claim 1, wherein the indication of the causal relationship comprises an indication of a type of the determined relationship on which the causal relationship is based.

8. The method of claim 1, wherein the experimental results data comprises a record of each of a plurality of experimental results described in corresponding publications.

9. The method of claim 1, wherein the experimental results data corresponds to results indicators relating to categories of the at least one performed experiment, the categories selected from the group consisting of positive manipulation, negative manipulation, non-intervention, and mediation.

10. The method of claim 1, further comprising:

receiving, from a user, additional experimental results data of an experiment, the experimental results data comprising indicators of (i) two variables, (ii) an experimental relationship indicator corresponding to two variables, and (iii) a result indicator corresponding to a type of the experiment.

11. A computer implementation system, comprising:

an input module that, by a processor, receives a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment;

a database containing experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target;

a map generating module that, from the database and by the processor, generates a research map based on the selection;

wherein the research map comprises nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges;

wherein one of the nodes represents the selected agent or the selected target;

wherein each of the edges comprises an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and

a hypothesis module that posits a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

12. The computer implementation system of claim 11, further comprising a display module that displays a graphical representation of the research map.

13. The computer implementation system of claim 11, wherein the indication of the causal relationship comprises a strength of the causal relationship.

14. The computer implementation system of claim 11, wherein the indication of the causal relationship comprises a positive or negative correlation between the connected nodes.

15. The computer implementation system of claim 11, wherein the indication of the causal relationship comprises an indication of a type of the determined relationship on which the causal relationship is based.

16. The computer implementation system of claim 11, wherein the experimental results data comprises a record of each of a plurality of experimental results described in corresponding publications.

17. The computer implementation system of claim 11, wherein the experimental results data corresponds to results indicators relating to categories of the at least one performed experiment, the categories selected from the group consisting of positive manipulation, negative manipulation, non-intervention and mediation.

18. A machine-readable medium comprising machine-readable instructions for causing a processor to execute a method comprising, comprising:

receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment;

by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target;

wherein the research map comprises nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges;

wherein one of the nodes represents the selected agent or the selected target;

wherein each of the edges comprises an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and

outputting, by a processor, a hypothesis positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.

19. A method comprising:

receiving, from a user, a selection of a parameter, the parameter representing a selected agent or a selected target of an experiment;

by a processor and from a database, generating a research map based on the selection, wherein the database contains experimental results data corresponding to relationships, determined by experiments, between (i) the agent or the target and (ii) another agent and/or target;

wherein the research map comprises nodes and edges, wherein each of the nodes represents an agent and/or a target and is connected to another node by one of the edges;

wherein one of the nodes represents the selected agent or the selected target;

wherein each of the edges comprises an indication of a causal relationship between connected nodes, the indication being based on one or more of the determined relationships; and

outputting, by a processor, a hypothesis map positing a causal relationship between two of the nodes A and C, separated by at least one other node B, wherein one of the nodes A, B, or C represents the selected agent or the selected target.



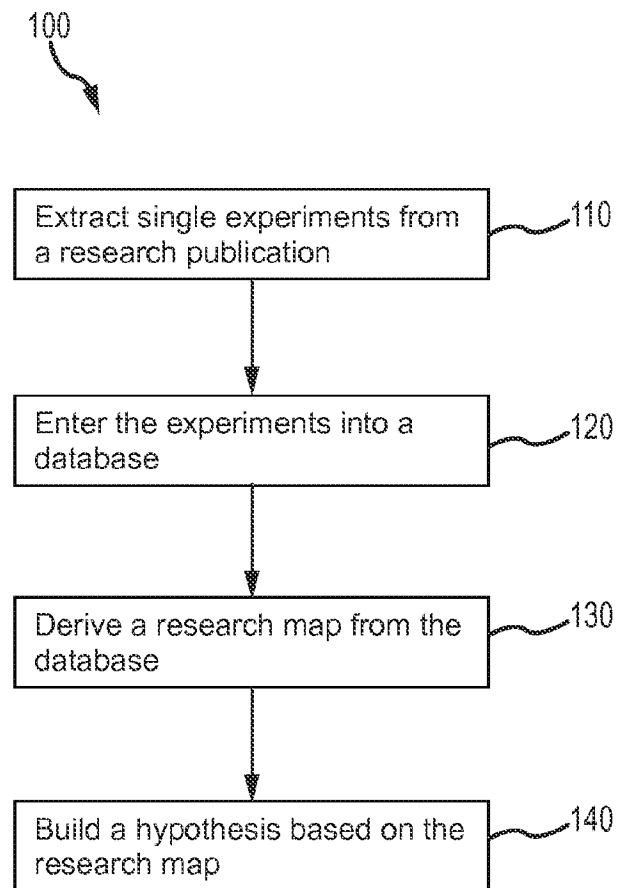


FIG.1

200



- 201 { Deletion of protein  $\alpha$  with genetic manipulation leads to deficits in synaptic plasticity  
in the hippocampal CA1 region
- 202 { Deletion of protein  $\alpha$  with genetic manipulation results in deficits in spatial learning
- 203 { Increases in Protein  $\alpha$  are observed in the earliest stages of spatial learning
- 204 { Increasing the expression of protein  $\alpha$  with a transgenic manipulation leads to  
enhancements in spatial learning
- 205 { Pharmacological block of protein  $\beta$  enhances synaptic plasticity in the hippocampal  
CA1 region
- 206 { Pharmacological block of protein  $\beta$  results in enhancements in in hippocampal  
spatial learning
- 207 { Pharmacological block of protein  $\beta$  results in increased activity of protein  $\alpha$   
in the hippocampal CA1 region
- 208 { Deletion of protein  $\beta$  with a genetic manipulation results in increased activity  
of protein  $\alpha$  in the hippocampal CA1 region

FIG.2

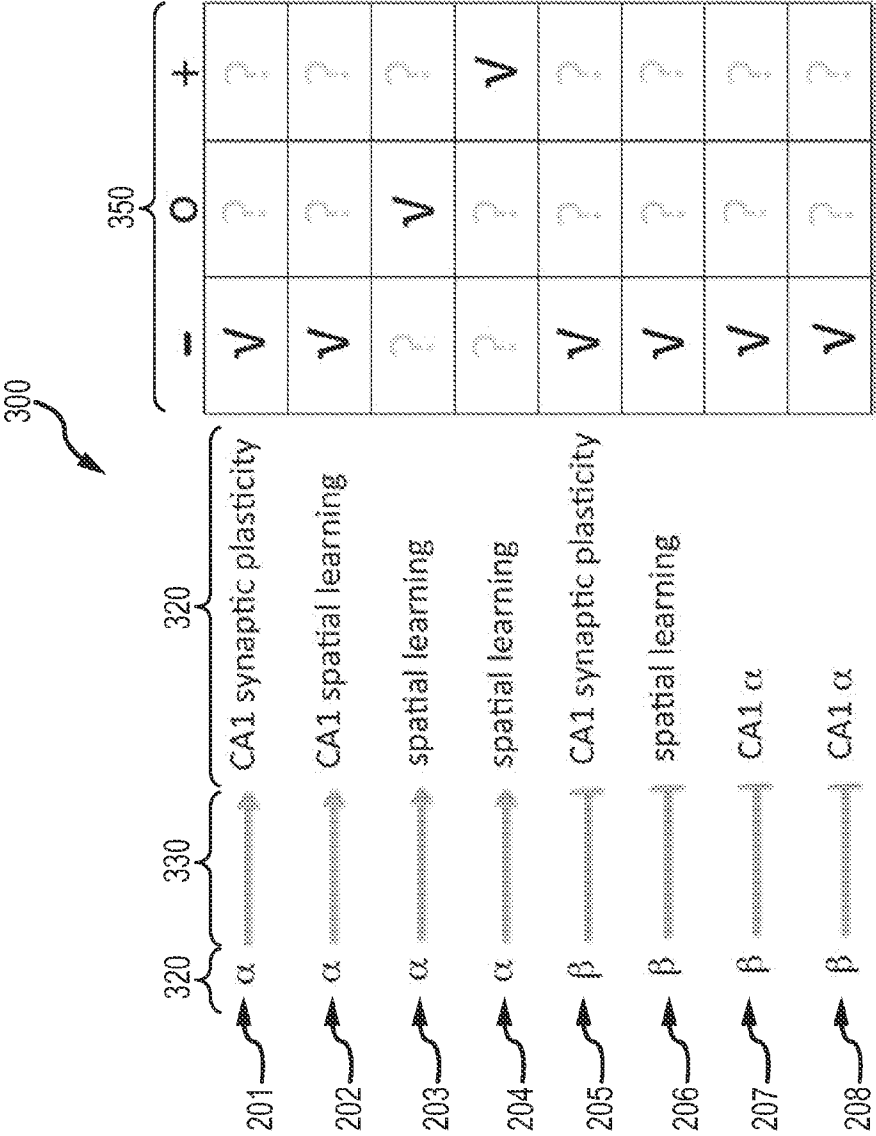


FIG.3

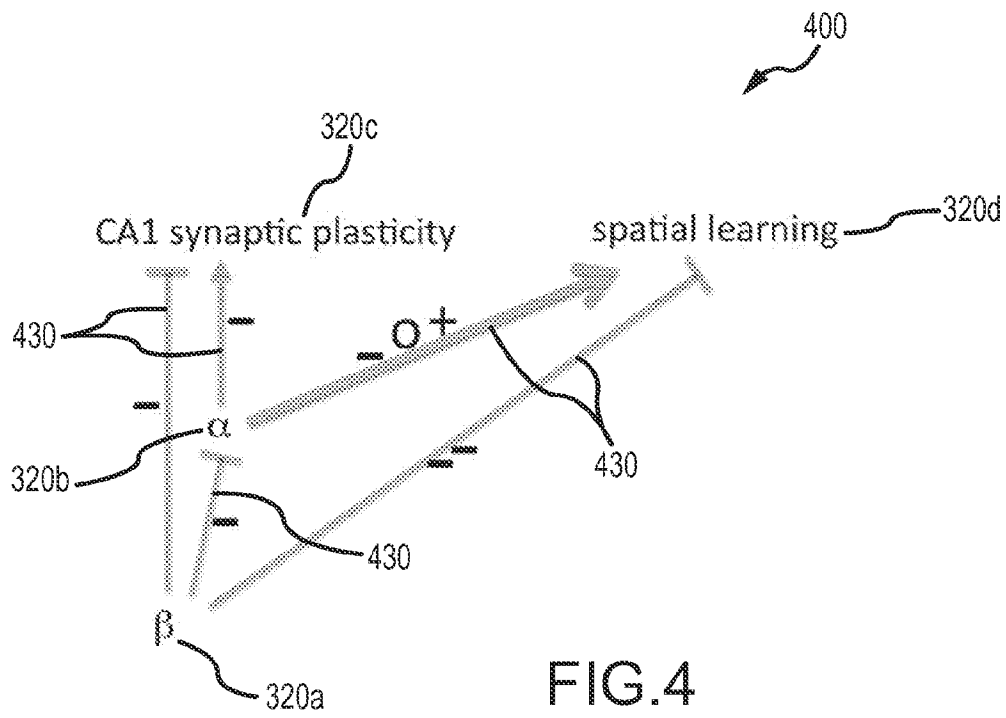


FIG.4

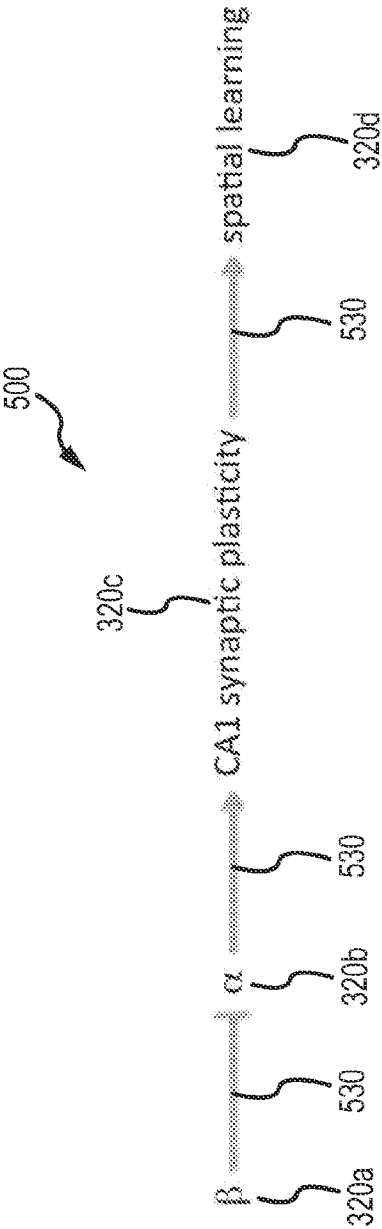


FIG.5

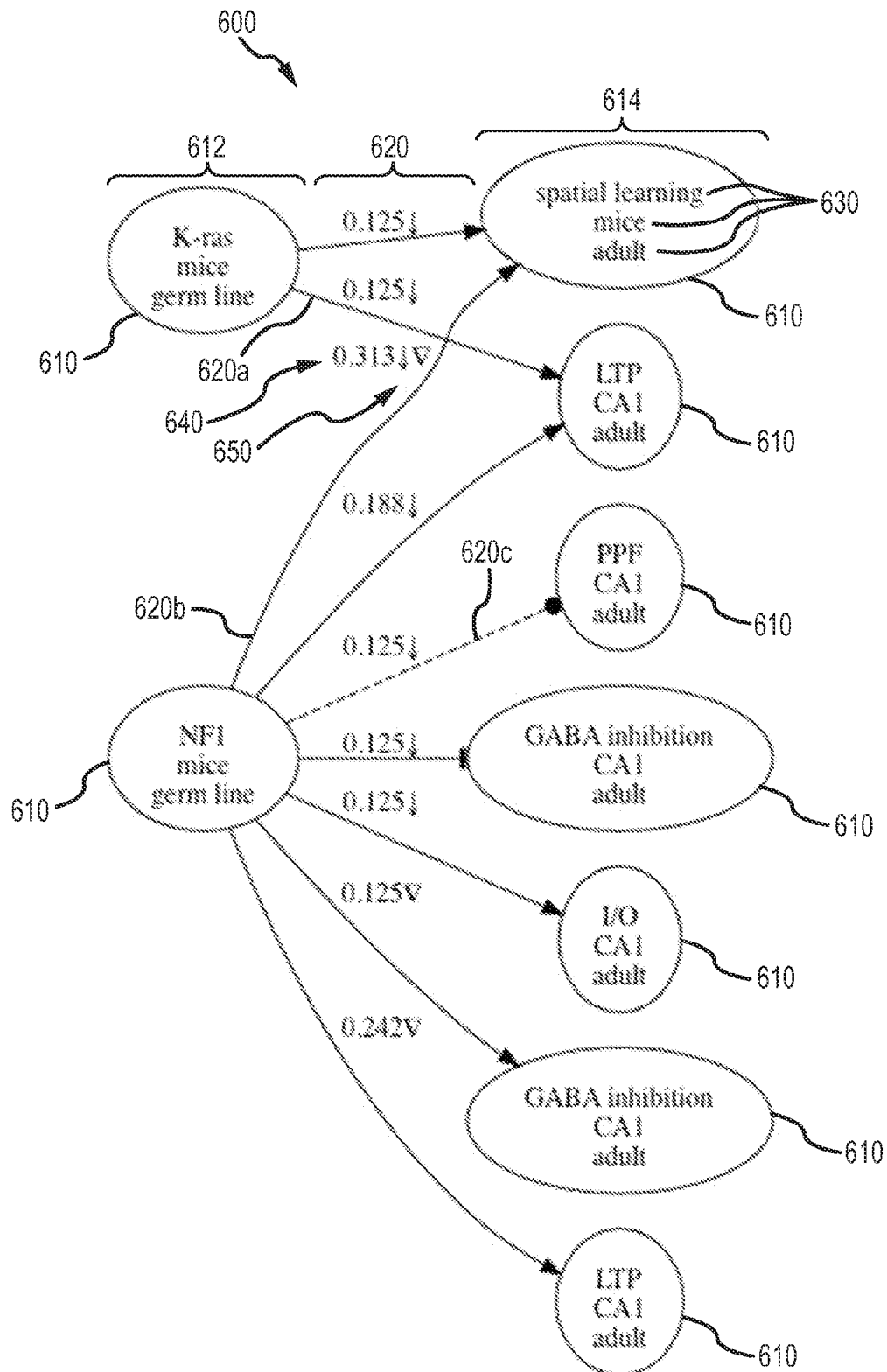


FIG. 6

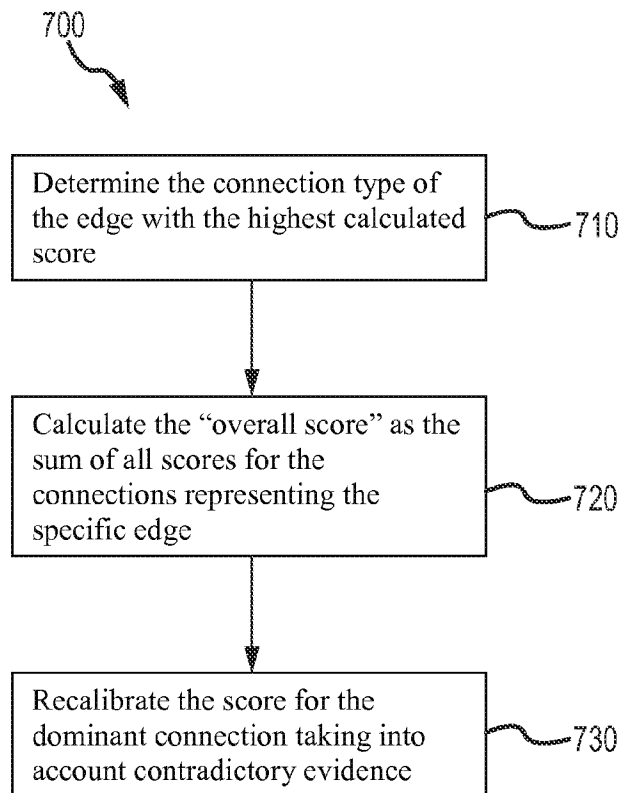


FIG.7

800

ARTICLE: NATURE 2002 JAN 31;415(6871):526-30 MECHANISM FOR THE LEARNING DEFICITS IN A MOUSE MODEL OF NEUROFIBROMATOSIS TYPE 1. COSTA RM, FEDEROV NB, KOGAN JH, MURPHY GG, STERN J, OHNO M, KUCHERLAPATI R, JACKS T, SILVA AJ DEPARTMENTS OF NEUROBIOLOGY, PSYCHIATRY AND PSYCHOLOGY, BRI, UNIVERSITY OF CALIFORNIA AT LOS ANGELES, LOS ANGELES, CALIFORNIA 90095-1761, USA. } 820

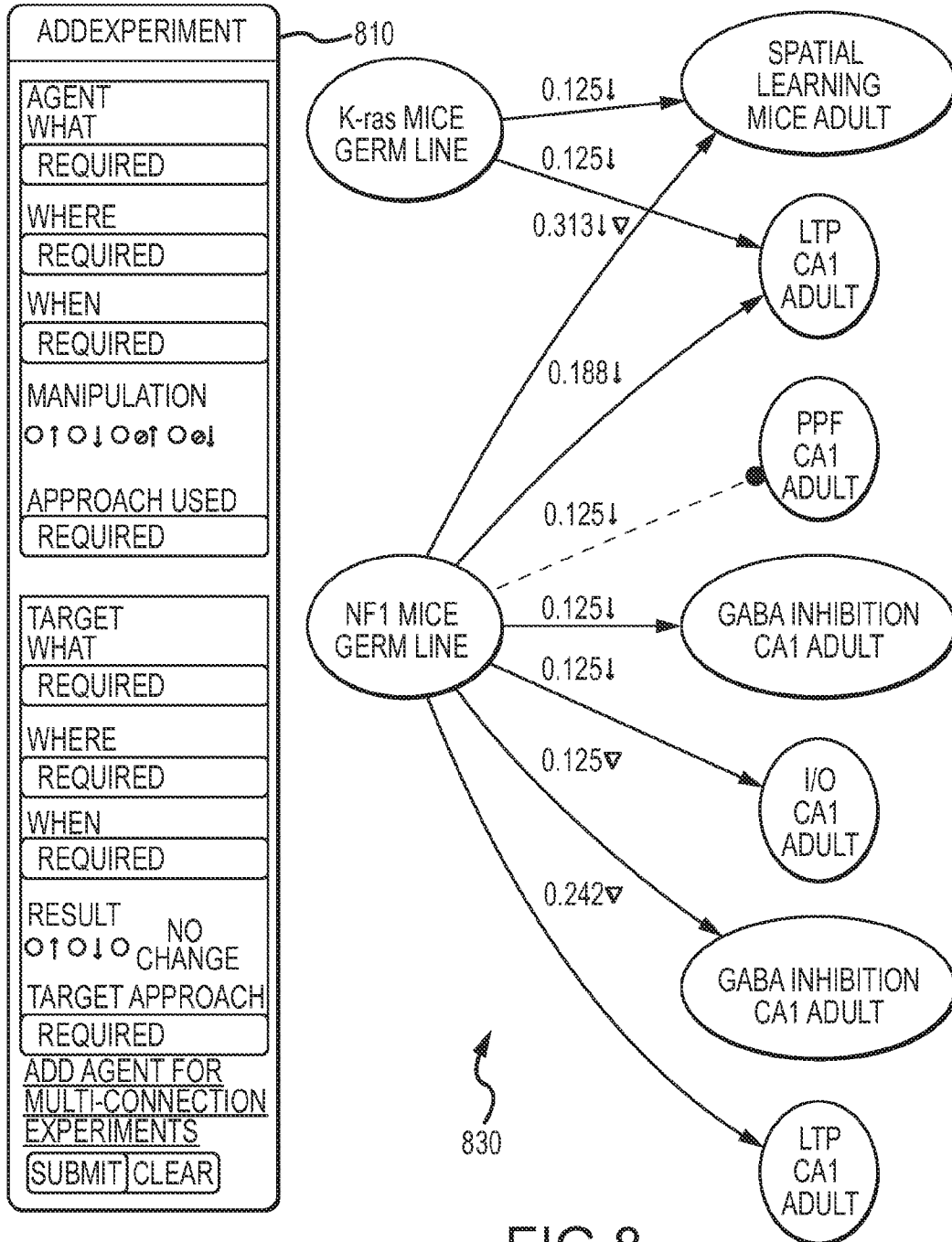
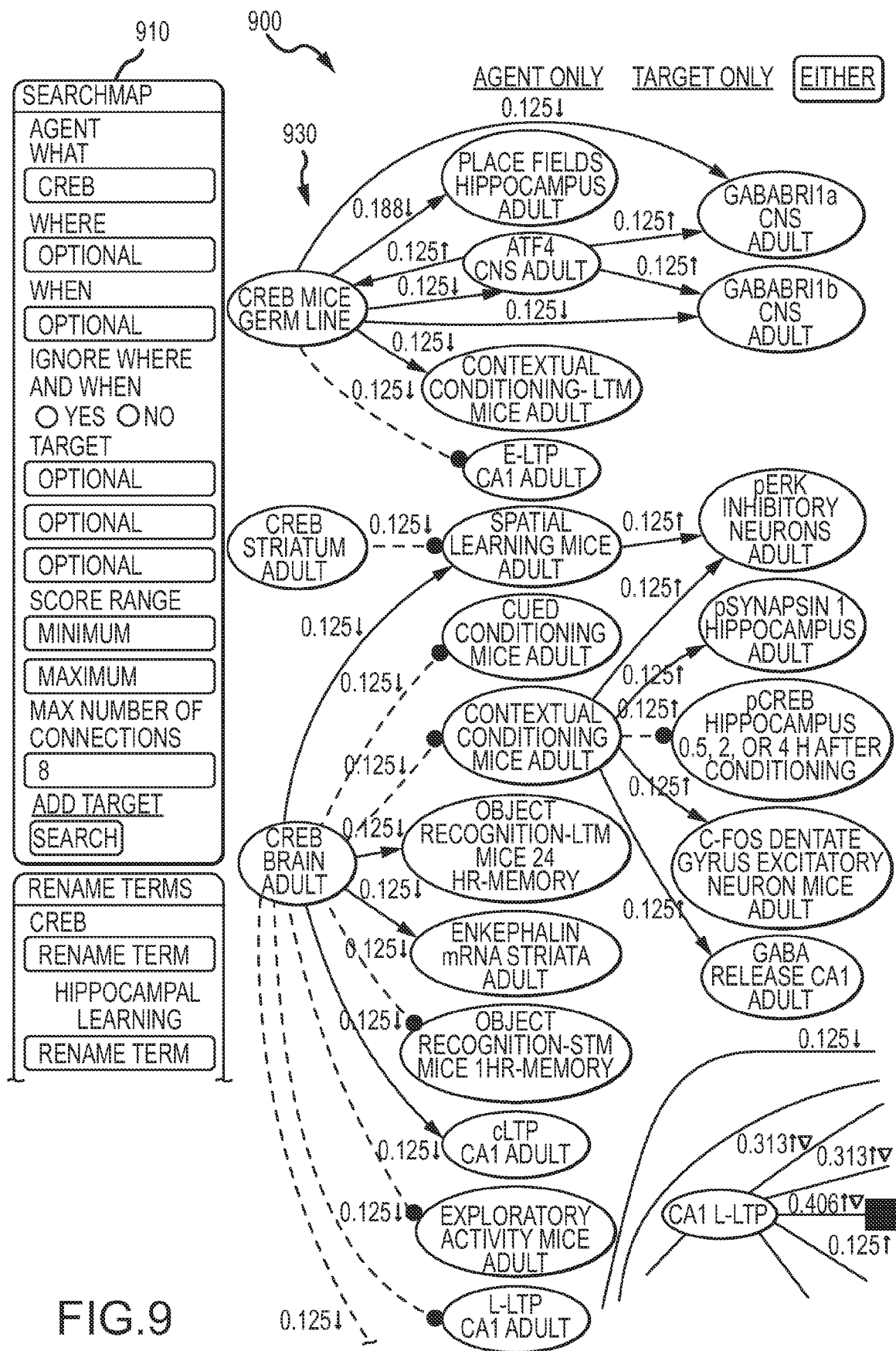
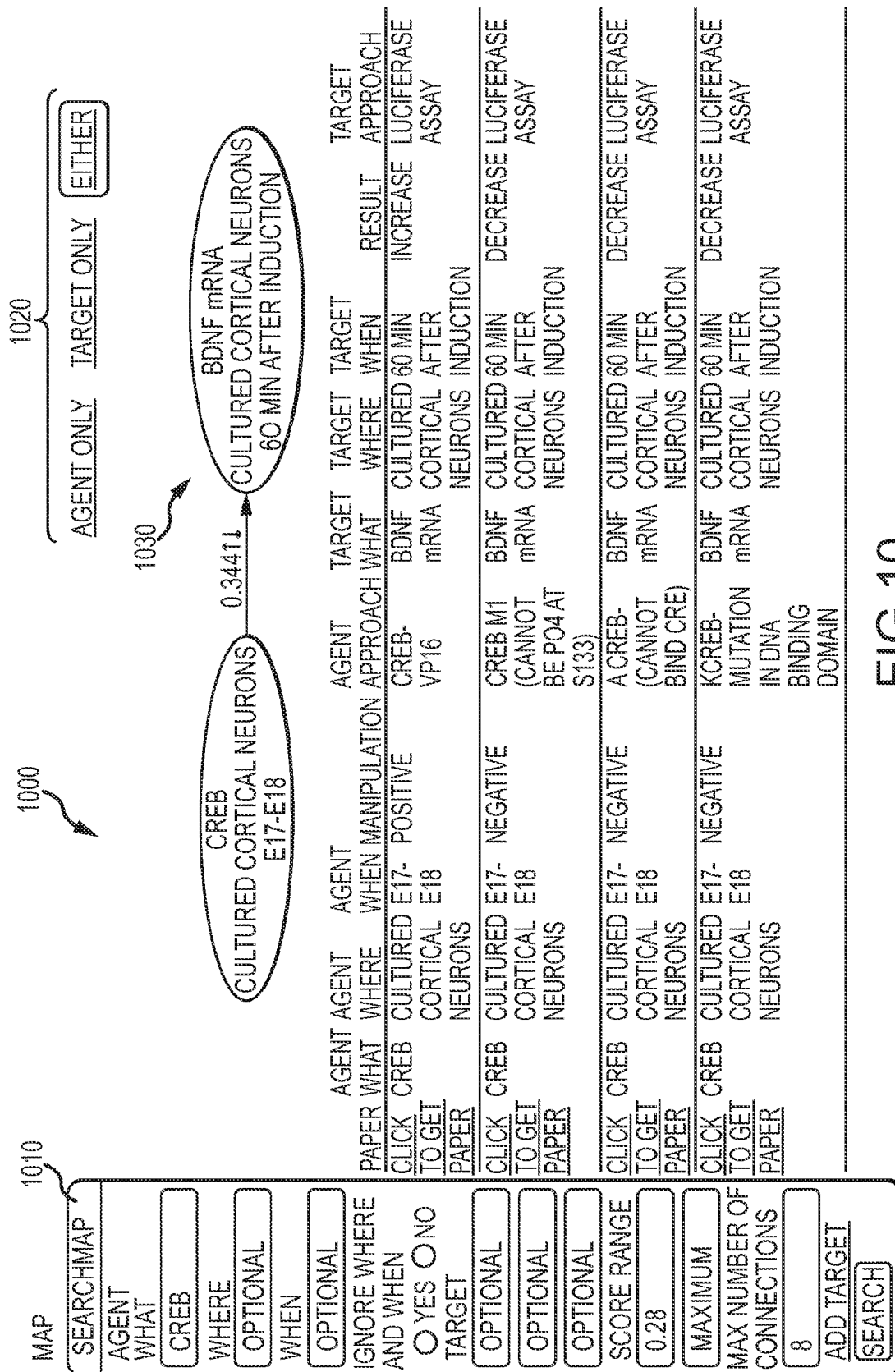


FIG.8







System 1100

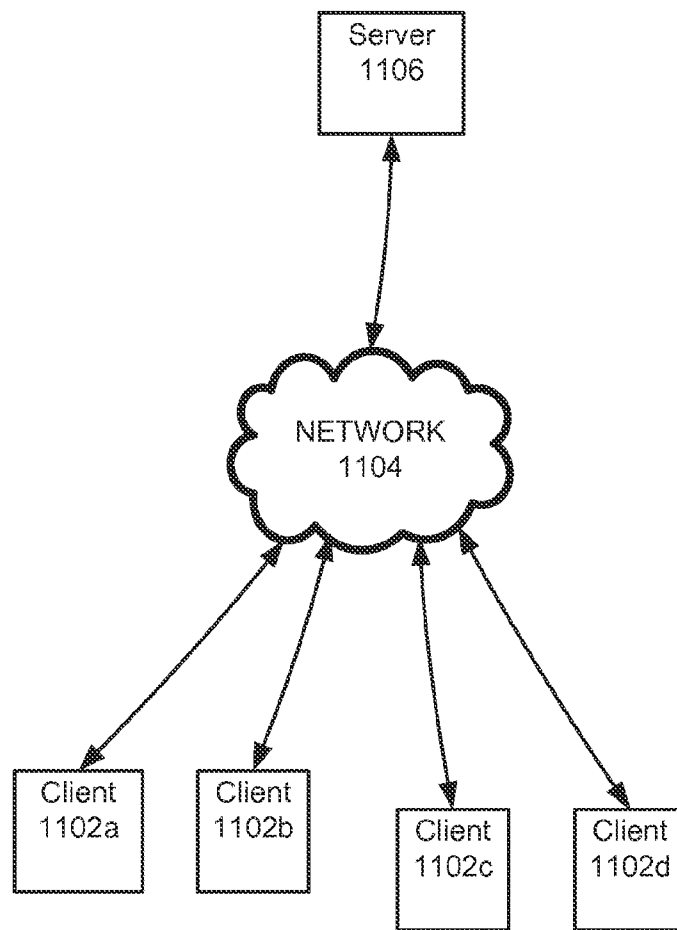


FIG. 11

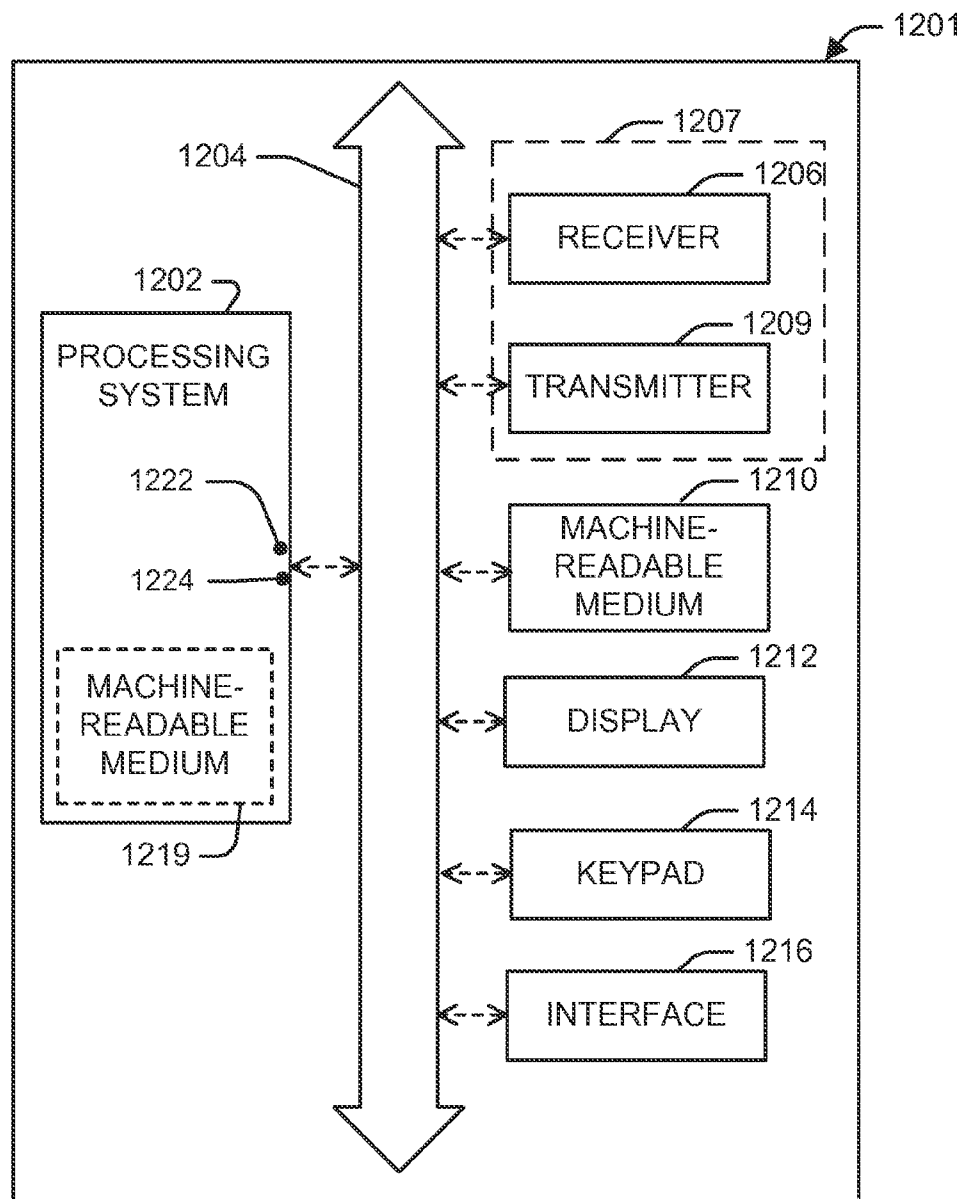


FIG. 12

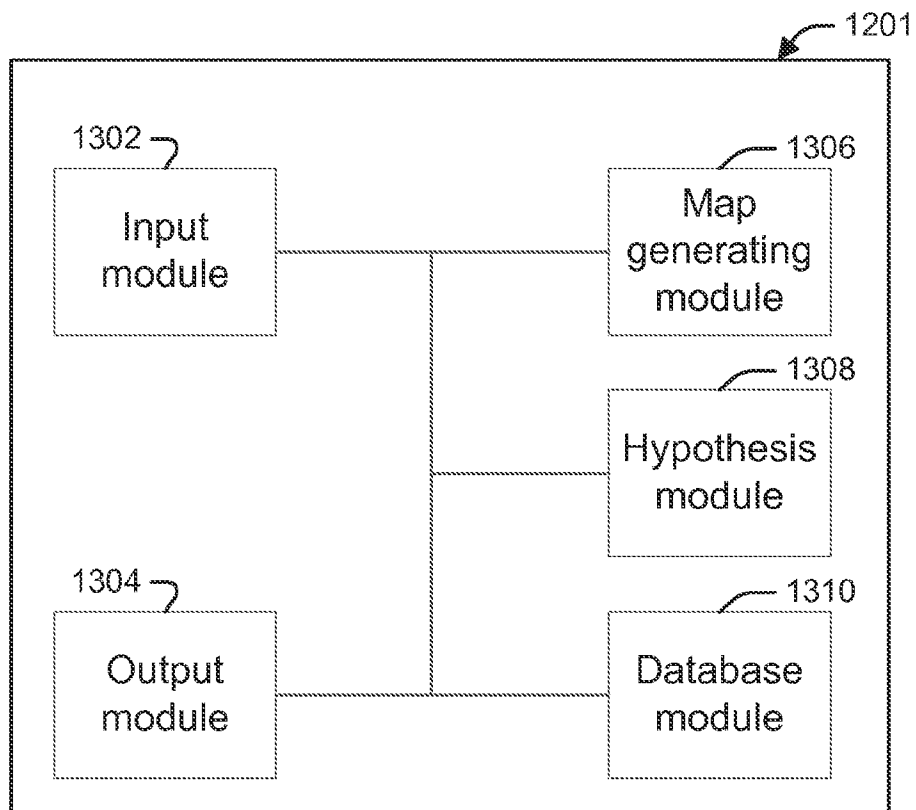


FIG. 13

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/049292

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 7/00 (2014.01)

CPC - G06Q 10/06 (2014.09)

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C12N 15/00; G01N 33/53; G06F 15/00, 17/00, 19/00, 7/00; G06N 5/00; G06Q 10/06; G06Q 30/00; H04B 15/00 (2014.01)

USPC - 702/179, 19, 32; 705/7.23, 7.29; 706/11, 45, 47, 52, 60; 707/679, 706, 737, 740, 758, 776, 798, 802, 999.001, E17.014

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - G06F 17/30321, 17/30333, 19/20; G06K 9/6224; G06N 5/022; G06Q 10/06, 10/063, 10 /0637, 10 /0639, 10/10, 10/105; G06Q 30/02, 30 /0201; G06Q 40/08; G06T 11/206 (2014.09)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Orbit, Google Patents, Google Scholar, Google.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2008/0288306 A1 (MACINTYRE et al) 20 November 2008 (20.11.2008) entire document	1-4, 7, 9, 11, 12, 15, 17-19 ----- 5, 6, 8, 10, 13, 14, 16
Y	US 6,952,688 B1 (GOLDMAN et al) 04 October 2005 (04.10.2005) entire document	5, 6, 8, 10, 13, 14, 16
A	US 8,392,418 B2 (BIRDWELL et al) 05 March 2013 (05.03.2013) entire document	1-19
A	US 8,250,107 B2 (ANDERSON) 21 August 2012 (21.08.2012) entire document	1-19
A, P	US 2014/0058782 A1 (GRAVES, JR.) 27 February 2014 (27.02.2014) entire document	1-19

☐ Further documents are listed in the continuation of Box C.

\* 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 another 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

11 October 2014

Date of mailing of the international search report

18 NOV 2014

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774