



(51) International Patent Classification:

G01N 33/18 (2006.01) B63G 8/00 (2006.01)
G06F 16/245 (2019.01) G01N 1/10 (2006.01)
G06F 16/248 (2019.01) G06F 17/00 (2019.01)

(21) International Application Number:

PCT/US2024/030978

(22) International Filing Date:

24 May 2024 (24.05.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/506,779 07 June 2023 (07.06.2023) US

(72) Inventor; and

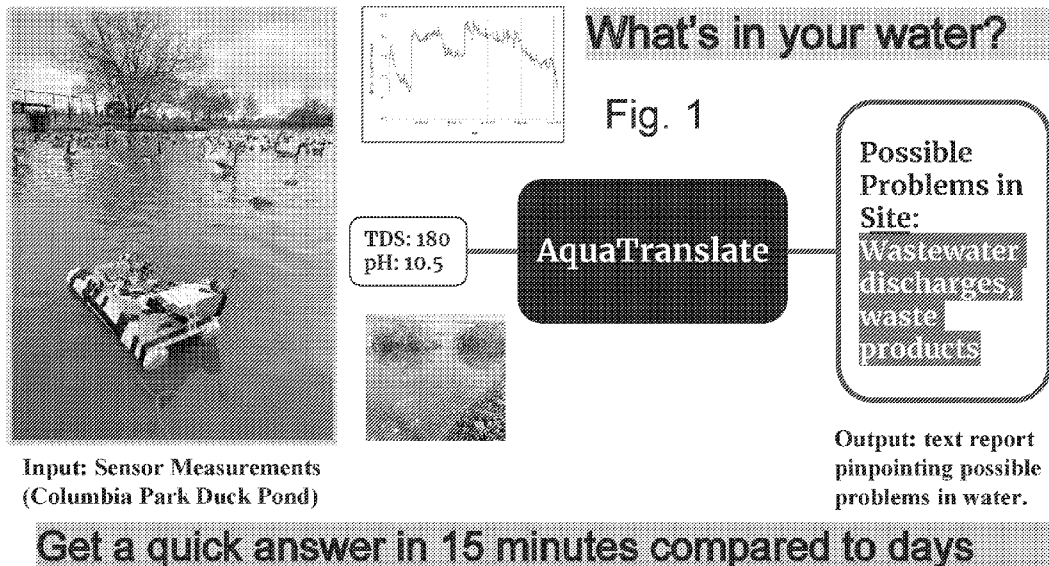
(71) Applicant: **CHOUDHURY, Prayona** [US/US]; 5716 W
11th Avenue, Kennewick, Washington 99338 (US).

(74) Agent: **ZIMOWSKI, Stephen**; c/o Gravis Law, 503
Knight Street, Suite A, Richland, Washington 99352 (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, CV, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,
KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH,
TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, CV,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,
SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

(54) Title: SYSTEMS AND METHODS FOR REAL-TIME TRANSLATION OF AQUATIC SENSOR MEASUREMENTS TO PROBABLE CAUSES VIA COMBINATION OF CAUSE-AND-EFFECT QUERIES AND LANGUAGE MODEL BASED SEMANTIC SIMILARITY



(57) Abstract: The invention provides a system for real-time and on-site analysis of water quality issues utilizing advanced algorithms, machine learning, and AI-based assessment, the system employing a state-of-the-art rover equipped with advanced sensors to detect issues and anomalies in an aquatic biome and supply collected data to computing units, also deployed locally on the rover, which then operate to assess the collected data by employing advanced algorithms to generate search engine queries, identify and extract pertinent terms from search results, assess likely causes of detected issues and anomalies, and provide real-time, on-site reports and analysis of findings.



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, ME, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *with international search report (Art. 21(3))*

-1-

TITLE OF THE INVENTION

SYSTEMS AND METHODS FOR REAL-TIME TRANSLATION OF AQUATIC SENSOR MEASUREMENTS TO PROBABLE CAUSES VIA COMBINATION OF CAUSE-AND-EFFECT QUERIES AND LANGUAGE MODEL BASED SEMANTIC SIMILARITY

FIELD OF THE INVENTION

[0001] The present invention relates to data collection and analysis in aquatic environments. More particularly, the present invention provides a system and device for acquiring location-based data on water quality, contaminants, and flora presence and providing data analysis and assessment of potential causes of water quality issues and anomalies shared across network devices in real time.

BACKGROUND OF THE INVENTION

[0002] Aquatic biomes, encompassing oceans, rivers, and lakes, play a vital role in maintaining the equilibrium of our planet's ecosystems. However, the increasing effects of climate change are causing rapid transformations in these environments, surpassing the natural adaptation capacities of resident organisms. Such changes have significant implications for aquatic ecosystems and industries reliant on them, such as agriculture and fishing. To ensure healthy crop growth and the sustenance of thriving fish populations, it is imperative to maintain optimal water quality.

[0003] Recognizing the critical importance of water quality, the present invention aims to revolutionize the process of data collection and analysis in aquatic environments. By leveraging cutting-edge technology, this system seeks to enhance the efficiency and accuracy of data processing, thereby aiding in the identification of potential problems and fostering environmental sustainability.

[0004] While certain systems currently exist to assist in this type of data accumulation and analysis, current systems suffer from several drawbacks, including the inability to assess potential causes of water quality issues in real time and at the data collection site. For example, Chinese Patent Publ. No. CN103175513B describes a system for monitoring hydrology and water quality of a river basin utilizing the

- 2 -

Internet of Things, including fixed and flowing sensors mounted in the river basin. Chinese Patent Publ. No. CN109470701A describes a system that performs biological monitoring of water treatment, including generating photos of pre-treated samples through the online monitoring device.

[0005] Neither of these systems, nor any others currently available for collecting and processing aquatic environmental data, utilize advanced data analysis techniques, such as cause-driven searching, vector embedding methods, or other sophisticated algorithms and AI-based analysis to assess and identify the potential causes of water quality issues and anomalies.

[0006] Accordingly, there remains a need in the art for a data collection and analysis system that can acquire location-based aquatic environment data, process, store, and analyze that data using advanced techniques, and provide system users with an assessment of potential causes of any water quality issues and anomalies, all in real time and on location.

SUMMARY OF THE INVENTION

[0007] It is therefore an object of the present invention to assist environmental and scientific analysis of water quality issues rapidly, on site, and utilizing advanced algorithms and machine learning to provide the most reliable and useful assessments to environmental analysts and agents in real time and at the location of data collection. It is a further object of the present invention to provide a system that learns and improves over time, utilizing machine learning and AI-based analysis to improve the reliability and thoroughness of its assessments, reports, and remedial recommendations.

[0008] The present invention thus provides an innovative Aquatic Data Analysis and Reporting System, referred to throughout as AquaTranslate, and comprises a state-of-the-art rover designed to collect comprehensive aquatic data. Equipped with advanced sensors, the rover is capable of measuring essential parameters such as temperature, pH, and total dissolved solids (TDS). These data points serve as crucial indicators of water quality.

- 3 -

[0009] One innovative aspect of AquaTranslate lies in its AI-based software, which undertakes the intricate task of analyzing the collected data. Leveraging sophisticated algorithms, the software generates comprehensive reports that elucidate the significance of the data points gathered. The system also employs machine learning techniques to identify and highlight potential issues within the water, such as the presence of algae or waste.

[0010] By streamlining the data collection and analysis process, AquaTranslate empowers researchers, environmentalists, and industries to proactively address water quality concerns. Rapid and accurate identification of problems allows for timely interventions, reducing the negative impact on aquatic ecosystems and the industries dependent on them.

[0011] AquaTranslate thus presents a novel system that addresses the urgent need for efficient aquatic data analysis. By facilitating improved understanding of water quality issues, this invention has the potential to contribute to environmental sustainability, safeguarding aquatic biomes and supporting industries vital to our global food supply.

[0012] Preferable embodiments of the present invention thus comprise a rover and specialized software designed to analyze and process data collected from aquatic environments. The rover is responsible for gathering various types of aquatic data, including but not limited to temperature, pH, and TDS (total dissolved solids) data. The present invention employs advanced algorithms to translate the collected data into comprehensive reports. Additionally, the system generates a list of potential issues or concerns related to the quality of the water being analyzed. The software component of the AquaTranslate System utilizes a combination of C++ programming language, specifically tailored for Arduino and sensor measurements, as well as Python programming language for AquaTranslate's natural language processing capabilities and querying Google search.

[0013] Preferable embodiments of the AquaTranslate system include a software system for real-time analysis of anomalous water quality that uses Cause-driven searches utilizing a search engine to identify potential causes of water quality

- 4 -

anomalies. A breadth-first search algorithm is used to generate many automated queries to a search engine (such as Google) that produce a list of search results. The system then summarizes search results using language model-generated vector-based comparisons that group items with high semantic similarity, referred to as vector embedding methods, to generate possible causes for anomalous water quality.

[0014] Preferable embodiments of the system then formulate queries to the web search engine, and the system groups potential causes by utilizing specific domain knowledge related to aquatic or marine environments. The inclusion of domain knowledge specific to aquatic or marine environments helps to enhance system performance and improve the reliability of the system's assessments and recommendations.

[0015] The present invention also preferably includes a hardware system enabling the real-time operation of the software system, which includes a low-cost, stable aquatic vehicle (also referred to as "rover" in text), preferably using a pontoon-boat structure with onboard electrical and computing units. The electrical units preferably include brushed motors that are used to propel "in-air" propellers that drive the rover. For rovers of such size, "in-air" motors and propellers require less power to propel the boat when compared to propellers mounted underwater, as underwater propellers face more hydrodynamic drag and require more expensive motors thereby increasing the cost of the rover.

[0016] The rover of the present invention is preferably equipped with at least two motors, and a user from land can preferably control the speed and direction of the vehicle using a radio controller that controls the operation of each of the motors. The rover also preferably includes one or more sensor units. The sensor units preferably consist of sensors to collect data on water temperature, pH, and total dissolved solids in addition to a microcontroller that collects GPS data. Each sensor unit preferably is attached to a separate Arduino microcontroller that executes a C++ module for collecting the sensor measurements.

[0017] A Raspberry Pi computer is also preferably mounted on-board the rover to communicate with the Arduino microcontrollers and execute a Python

- 5 -

software that translates sensor measurements into natural language based description of potential causes. The Raspberry Pi system is preferably remotely controllable (using a messaging software framework) such that a user from land can send a control message to the on-board software in the waterborne rover to execute the AquaTranslate software system.

[0018] Upon generation of probable causes creating the environmental factors detected by the sensor measurements, the AquaTranslate system operating in the aquatic environment (such as river) notifies the user standing on land (such as riverbank) via a text message over bluetooth or cellular interface. The system's analysis and messaging preferably include an assessment of the most likely cause(s) of the environmental factors and, in some preferable embodiments, remedial actions that can be taken immediately or over the long term to resolve the water quality issues or anomalies.

[0019] The one or more computing units utilized by the AquaTranslate system, such as Arduino microcontrollers and Raspberry Pi, are specifically selected and designed to have minimal volume and weight, enabling attachment to an autonomous or human-guided rover operating in aquatic environments. Such design increases the floating stability of the rover and minimizes the rover's on-board power requirements extending the operating range of the rover and the data collection and analysis system as a whole.

[0020] As those of skill in the art will appreciate, the present invention is not limited to the embodiments and arrangements described above. Other objects of the present invention and its particular features, arrangements, and advantages will become more apparent from consideration of the following brief description of the drawings and detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] FIG. 1 depicts a rover and sample data analysis and schematic of the AquaTranslate system according to preferable embodiments of the present invention.

[0022] FIG. 2 depicts various design options for a rover according to the

preferable embodiments of the present invention depicted in Fig. 1.

[0023] FIG. 3 depicts features of a rover according to the preferable embodiments of the present invention depicted in Figs. 1-2.

[0024] FIG. 4 depicts features of a rover according to the preferable embodiments of the present invention depicted in Figs. 1-3.

[0025] FIG. 5 depicts a schematic of a circuit diagram for the features of the rover according to the preferable embodiments of the present invention depicted in Figs. 1-4.

[0026] FIG. 6 depicts a data flow schematic for the transmission and processing of acquired data according to the preferable embodiments of the present invention depicted in Figs. 1-5.

[0027] FIG. 7 depicts computing unit features utilized with the AquaTranslate system according to the preferable embodiments of the present invention depicted in Figs. 1-6.

[0028] FIG. 8 depicts a schematic diagram of the process of converting acquired data to output report according to the preferable embodiments of the present invention depicted in Figs. 1-7.

[0029] FIG. 9 depicts a schematic illustration of the AquaTranslate system's use of querying through a search engine according to the preferable embodiments of the present invention depicted in Figs. 1-8.

[0030] FIG. 10 depicts a schematic illustration of the AquaTranslate system's use of querying through a search engine according to the preferable embodiments of the present invention depicted in Figs. 1-9.

[0031] FIG. 11 depicts a schematic illustration of the AquaTranslate system's extraction of applicable terms identified through repeated search engine queries according to the preferable embodiments of the present invention depicted in Figs. 1-10.

[0032] FIG. 12 depicts a schematic illustration of the AquaTranslate system's

- 7 -

extraction of common factors using the principle of semantic similarity obtained by querying language models according to the preferable embodiments of the present invention depicted in Figs. 1-11.

[0033] FIG. 13 depicts a schematic illustration of the AquaTranslate system's extraction of common factors using the principle of semantic similarity obtained by querying language models according to the preferable embodiments of the present invention depicted in Figs. 1-12.

[0034] FIG. 14 depicts a exemplary report produced by the AquaTranslate system based upon data acquired and processed from one or more aquatic biomes according to the preferable embodiments of the present invention depicted in Figs. 1-13.

[0035] FIG. 15 depicts an overhead view of an area selected for exemplary assessment by the AquaTranslate system according to the preferable embodiments of the present invention depicted in Figs. 1-14.

[0036] FIG. 16 depicts an exemplary data collection and analysis as provided by the AquaTranslate system according to the preferable embodiments of the present invention depicted in Figs. 1-15.

DETAILED DESCRIPTION OF THE INVENTION

[0037] The following detailed description illustrates and describes the present invention by way of example, not by way of limitation of the principles of the invention. This description will enable one skilled in the art to make and use the invention and describes several embodiments, features, arrangements, adaptations, variations, alternatives, and uses of the invention, including what is presently believed to be the best mode of carrying out the invention. Those of skill in the art will appreciate other variations and arrangements possible without straying beyond the principles of the invention, and the present invention is not limited to those embodiments and arrangements described hereafter.

[0038] Preferable embodiments of the present invention utilize two microcontrollers or sensor units to collect data, which is sent to a Raspberry Pi. The Pi

- 8 -

receives the numerical TDS (total dissolved solids) and pH data approximately 20 seconds after the rover is in the water. This data is provided to the AquaTranslate software system. The software system categorizes the numerical data into a set of strings. Next, a starter search is made using the string previously categorized. String variables such as “high pH” are replaced in a certain format that all the automated searches follow. The format that all AquaTranslate search engine queries follow is “What causes” + the problem such as “high pH” + “in” + the location, such as “rivers”. This all adds up creating the search engine query “What causes high pH in rivers”. Then AquaTranslate uses a web-search API, such as Google, Bing, Yahoo, or the like, to submit the search engine query, parse the HTML response and save the answer.

[0039] A designed algorithm is then preferably used to take the previous query result and identify the main “branching variables” from the results. For example, the results from the query “what causes high pH in rivers” is “In general, chemicals, minerals, pollutants, soil or bedrock composition, and any other contaminants that interact with a water supply will create an imbalance in the water's natural pH of 7.” Preferable embodiments of the AquaTranslate code use Spacy, a natural language processing library, which uses machine learning to extract the nouns such as “chemicals”, “minerals”, “pollutants”, “soil”, and “bedrock composition.” The algorithm will then input these variables into their own search such as “what causes chemicals in rivers”. The same variables are preferably stored in a list for possible problems.

[0040] Preferable embodiments of the algorithm will iterate a maximum of five times but will stop when either the results are quite similar or the search engine does not provide a useful answer. Finally, AquaTranslate will create and provide a report compiling the sequence of searches for their results, with a list containing terms that are possible aquatic problems. This process is preferably repeated twice to provide a list of possible problem terms for both TDS and pH.

[0041] Preferable embodiments of the AquaTranslate system and invention finish the process by cross-analyzing both lists to identify similar terms in both lists. This task is preferably performed using OpenAI's davinci language model. The search

- 9 -

results obtained in the previous step are converted on-the-fly into a query, and the query is sent to OpenAI language model server, which groups similar terms into separate lists based on their semantic similarity. The three lists containing the most terms, along with the numerical data are texted back to the user over bluetooth or cellular network, all while the AquaTranslate rover is in the water.

[0042] Certain exemplary embodiments of the present invention have been described herein in considerable detail to provide those skilled in the art with the information needed to apply the novel principles of the invention and to construct and use such exemplary and specialized components as are required. While the present invention has been described with reference to particular embodiments and arrangements of parts, features, and the like, it is not limited to these embodiments or arrangements. Indeed, modifications and variations will be ascertainable to those of skill in the art, all of which are to be understood as within the spirit and scope of the present invention.

- 10 -

What is claimed is:

1. An aquatic data analysis system comprising:

a waterborne craft navigable through an aquatic environment, the waterborne craft comprising one or more sensors for acquiring data on the aquatic environment;

an aquatic environment analysis module comprising one or more algorithms for analyzing anomalies in the aquatic environment by processing the data on the aquatic environment acquired by the one or more sensors of the waterborne craft, assessing the associated water and other aquatic material, and identifying anomalies found therein;

an inquiry module comprising one or more algorithms for generating one or more queries to identify potential causes of the anomalies identified by the aquatic environment analysis module, running the one or more queries on one or more internet search engines, and accumulating the results of the one or more queries; and

an assessment module comprising one or more algorithms for summarizing the results of the one or more queries using language model generated vector-based comparisons, grouping the potential causes of the anomalies utilizing accumulated system knowledge specific to aquatic environments into one or more causal groups, and generating an assessment of the one or more causal groups;

wherein the one or more sensors of the waterborne craft are in electronic communication with at least one of the aquatic environment analysis module, the inquiry module, and the assessment module enabling real-time operation of the aquatic data analysis system as the waterborne craft navigates throughout the aquatic environment.

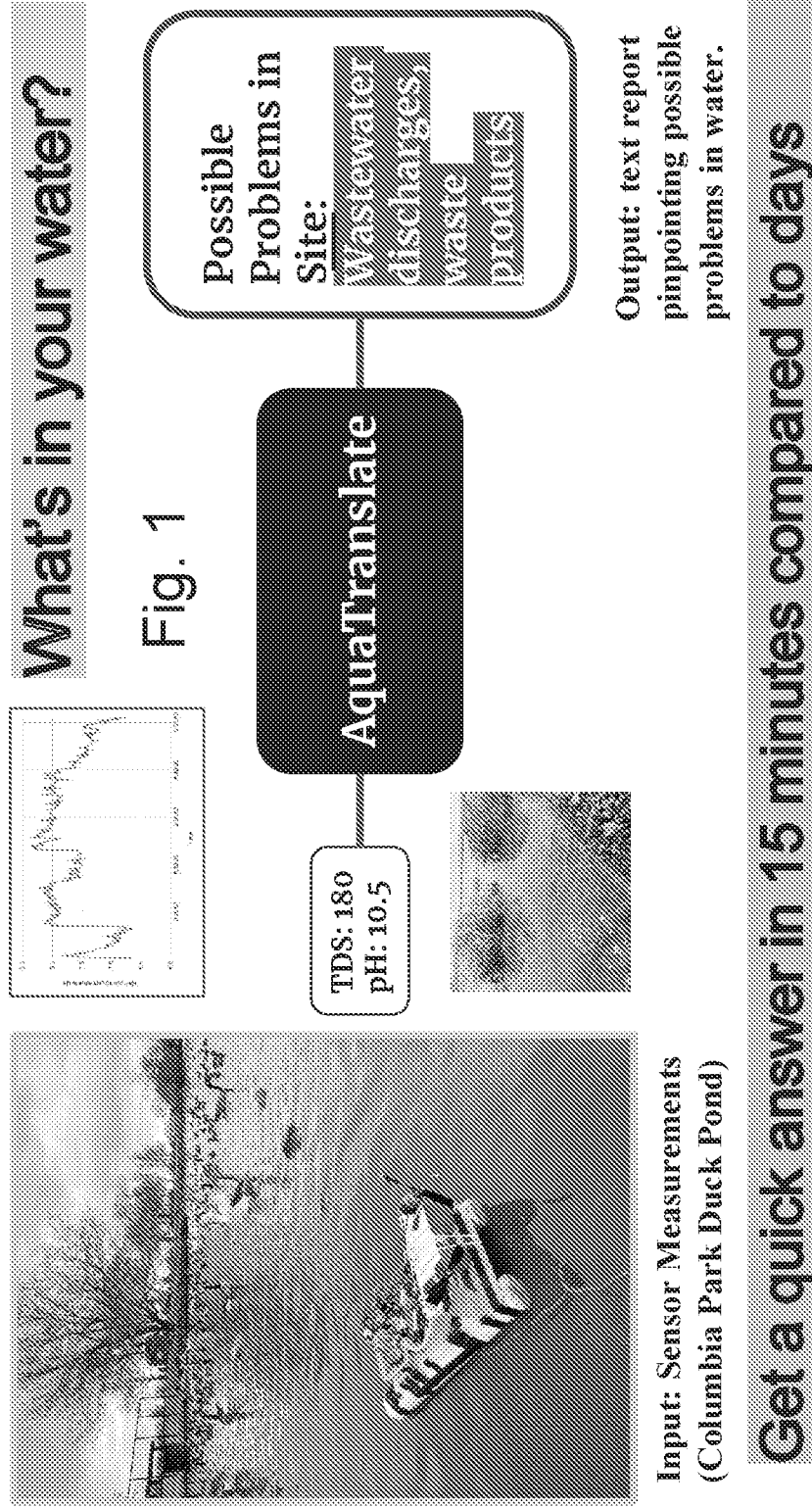
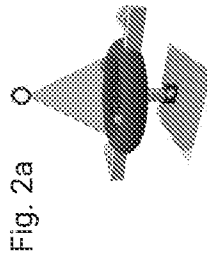
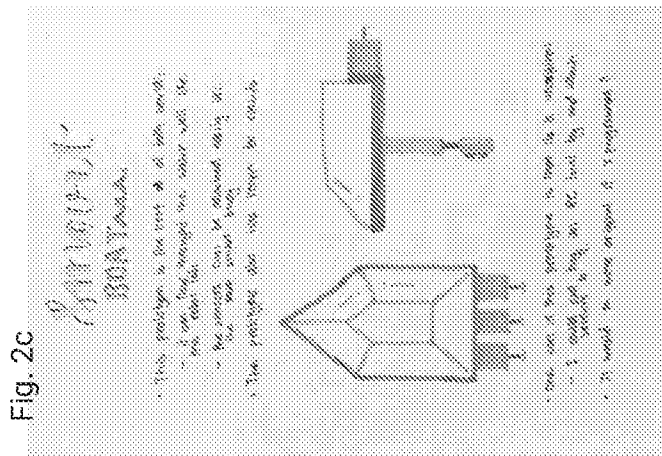


Fig. 2: Considerations for Novel Design

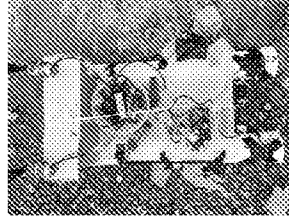


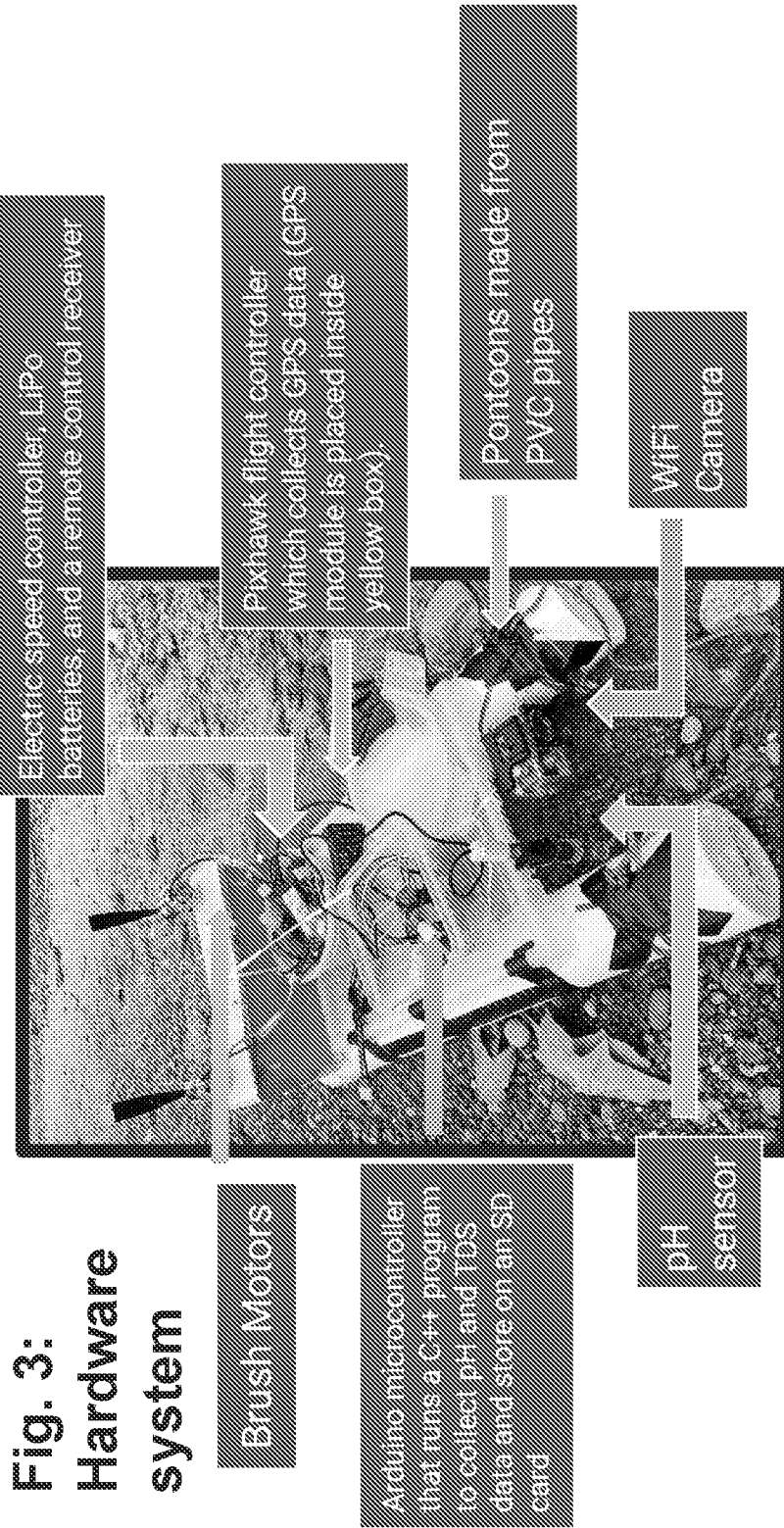
When choosing the design of my vehicle, I had three main options - 1) building a motorized buoy, 2) a robotic fish, or 3) a boat design. However, each design had their own pros and cons, so in the end, my final design takes elements from each option.

Design options



I was inspired by the floatability of the buoys (made me select pontoons), motor propellers from the fish and overall idea of steering a boat with mounted sensors





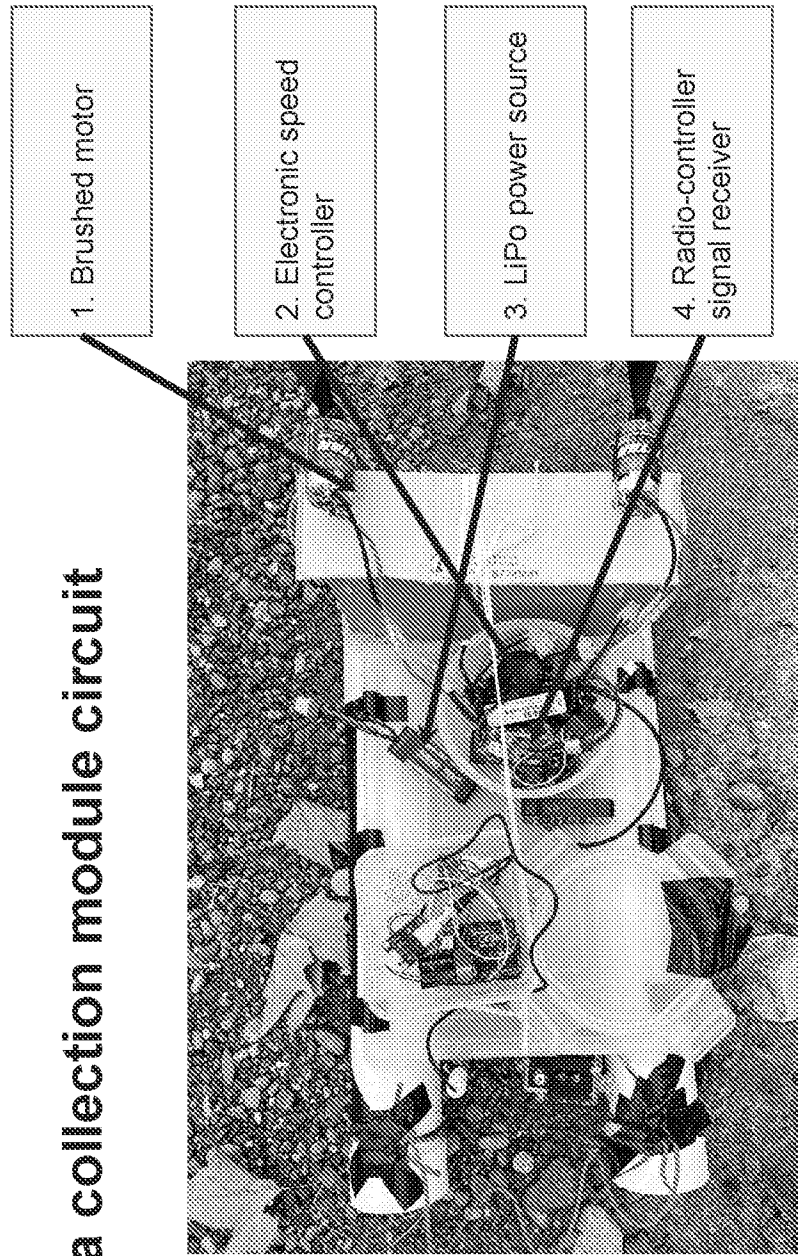


Fig. 4: Data collection module circuit

Fig. 5: Circuit Diagram

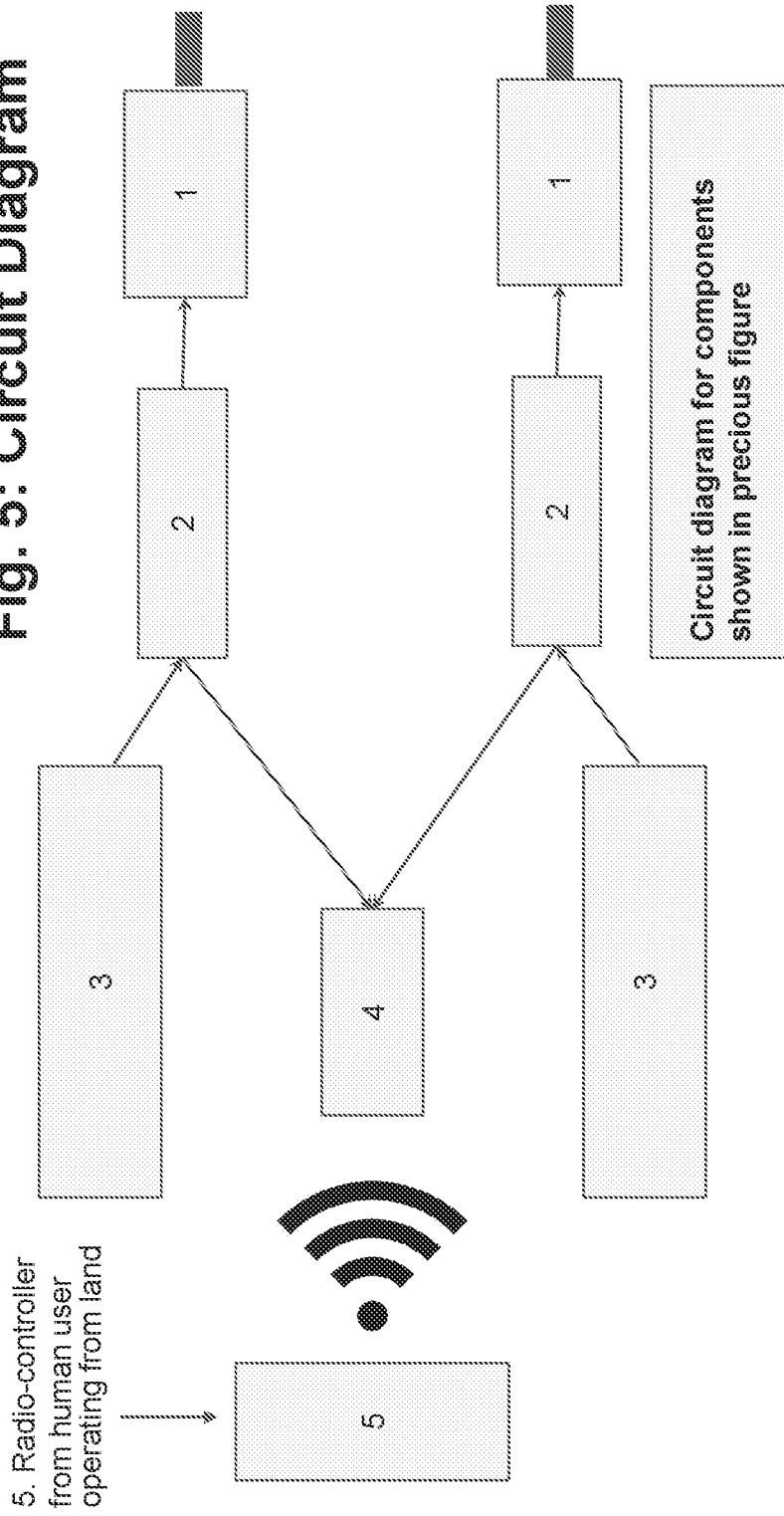


Fig. 6: Data flow: From sensors to AquaTranslate input

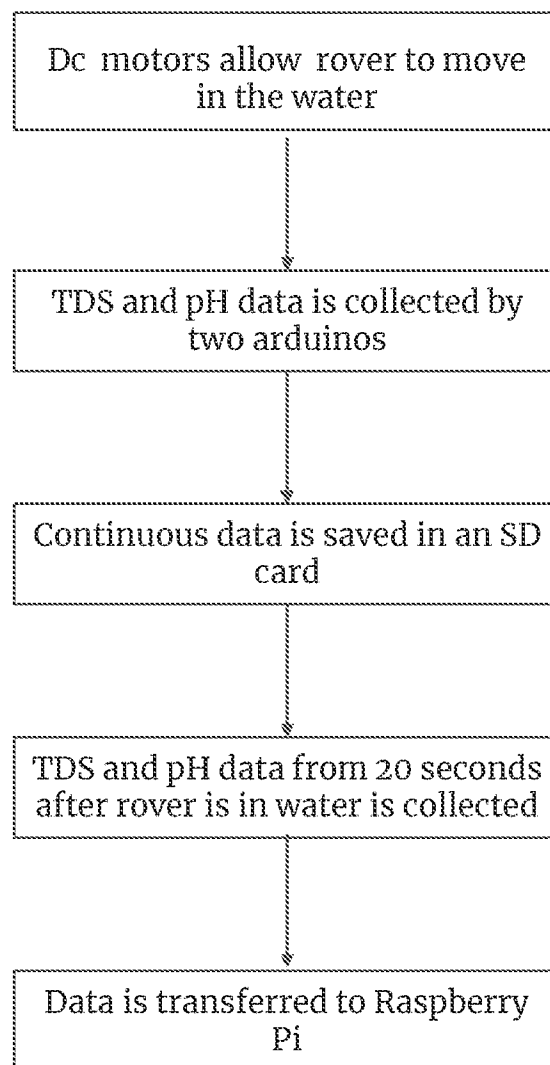
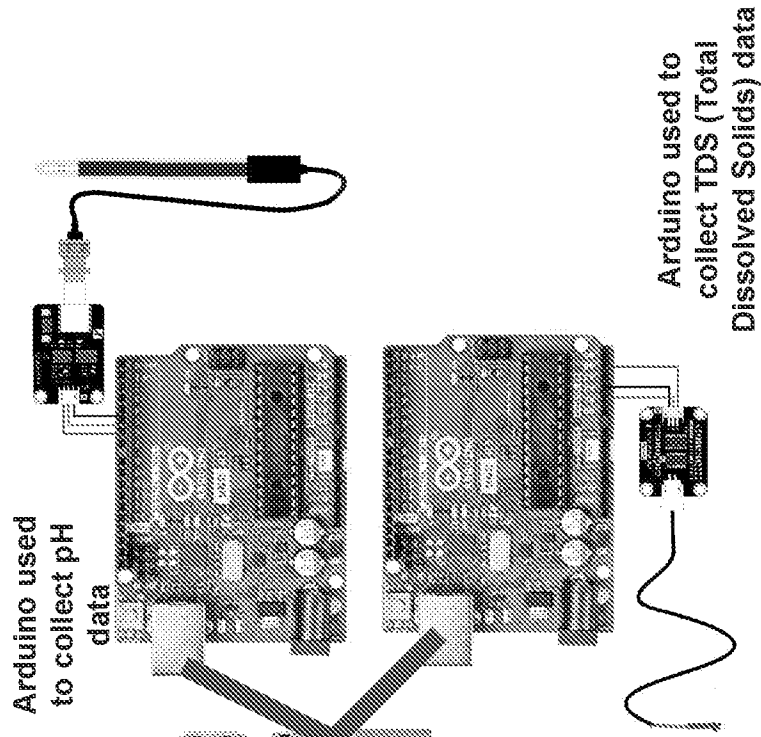


Fig. 7: Hardware connection diagram from sensors to computing unit



Raspberry Pi used to:

1. Receive pH and TDS data
2. Input the data into Aqua Translate Software
3. Text answers back to user

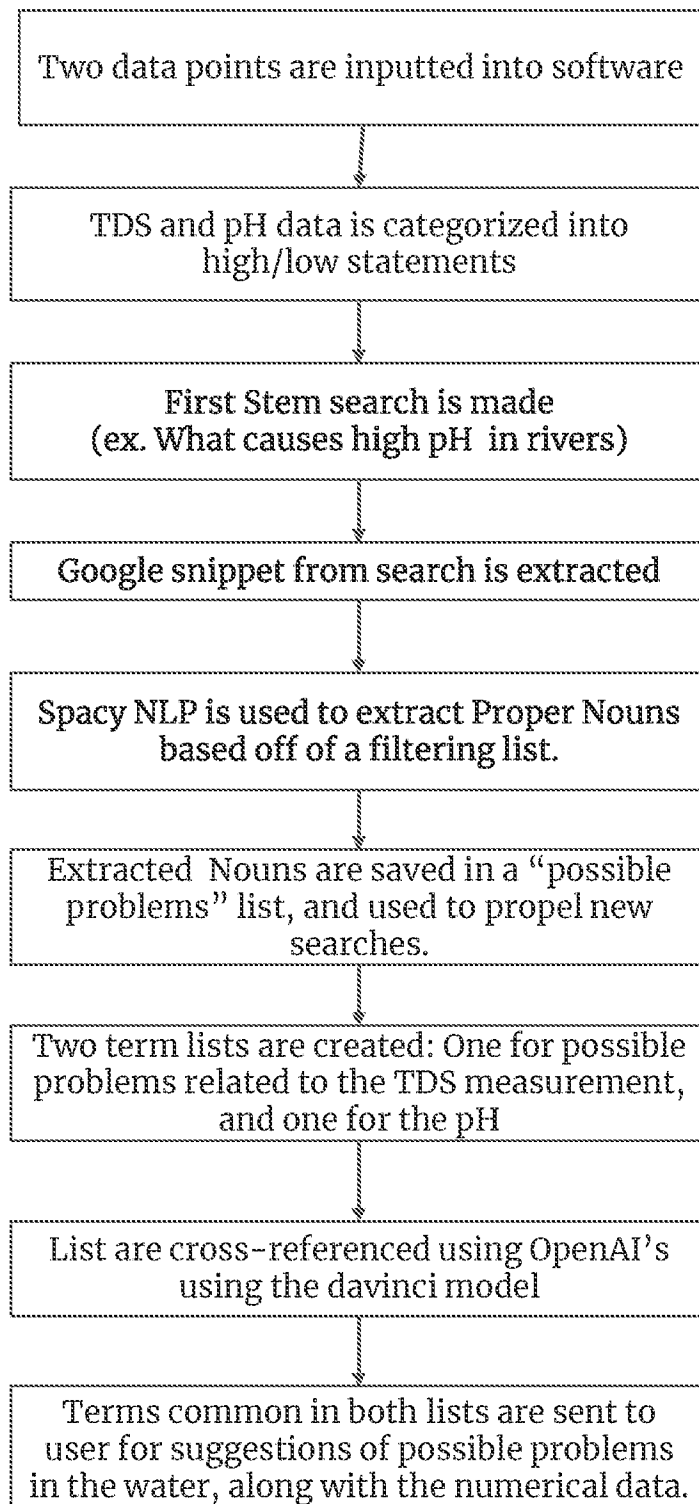
Fig. 8: Conversion from raw sensor data to output

Fig. 9: Illustration of automated querying of a search engine

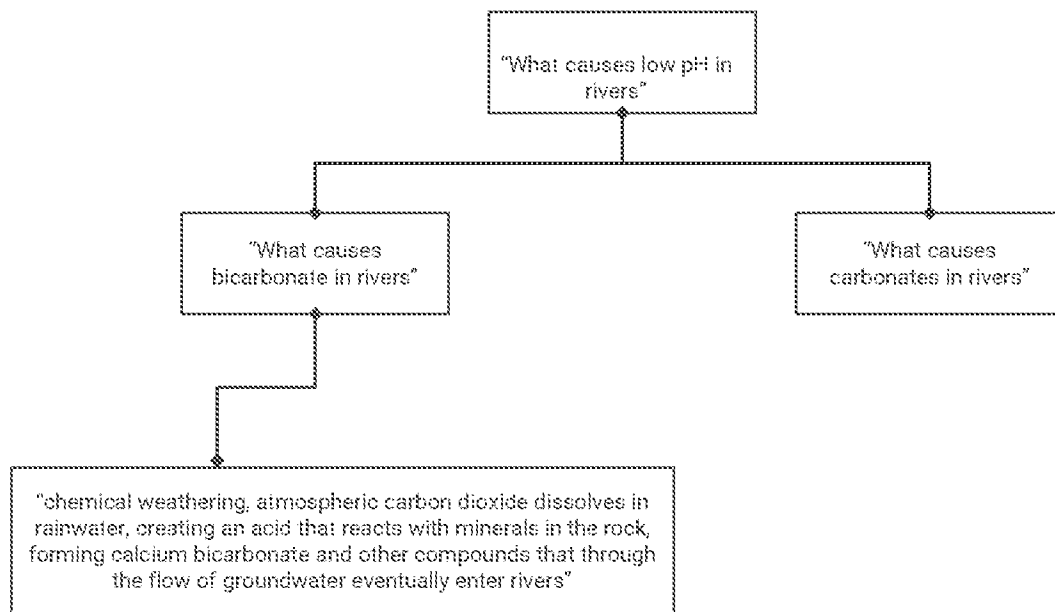


Fig. 10: Illustration of automated querying of a search engine (additional example)

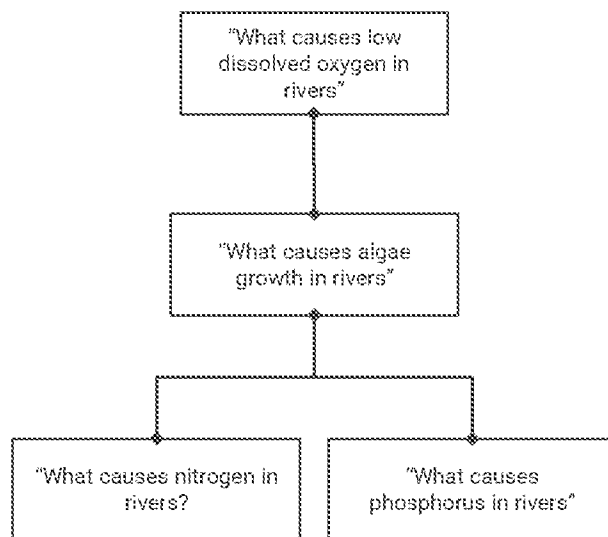


Fig. 11: Illustration of terms extracted from repeated queries

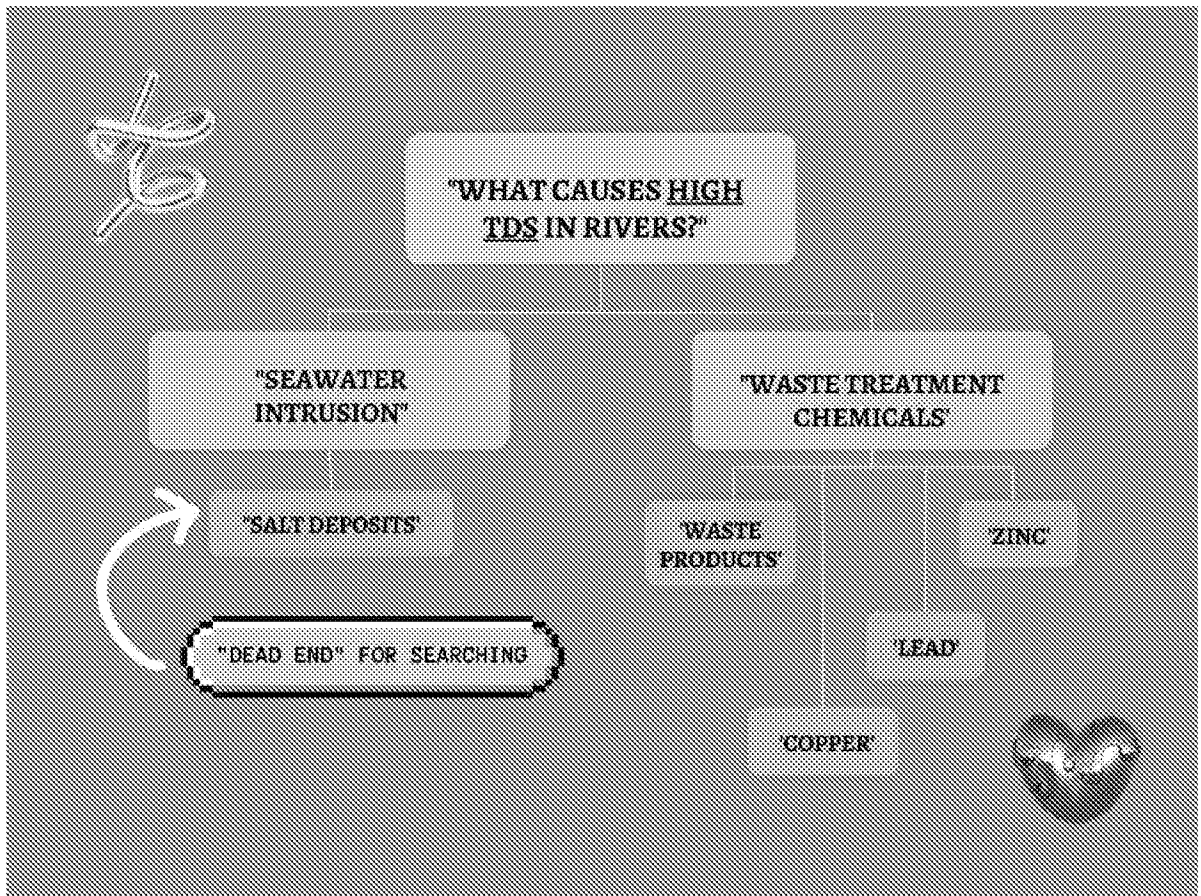


Fig. 12: Extracting common factors using the principle of semantic similarity and obtained via querying a language models (example 1)

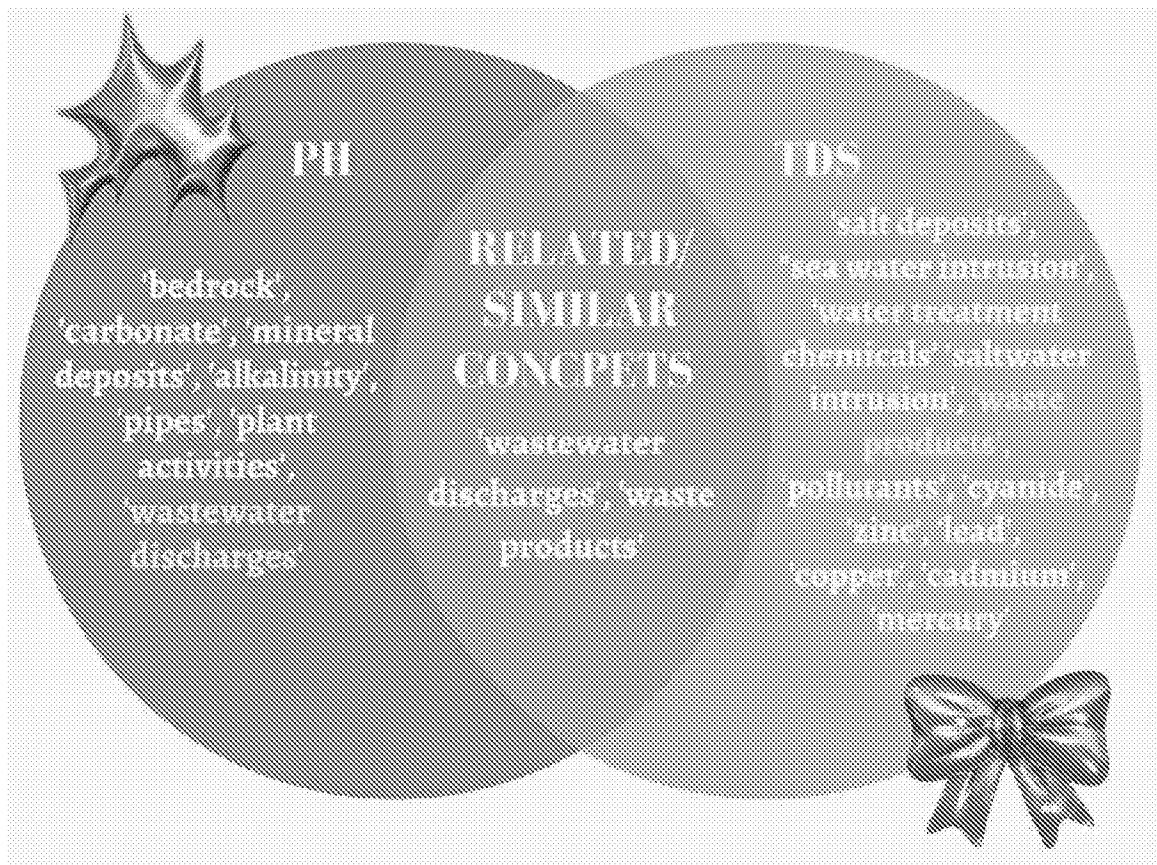


Fig. 13: Extracting common factors using the principle of semantic similarity and obtained via querying a language models (example 2)

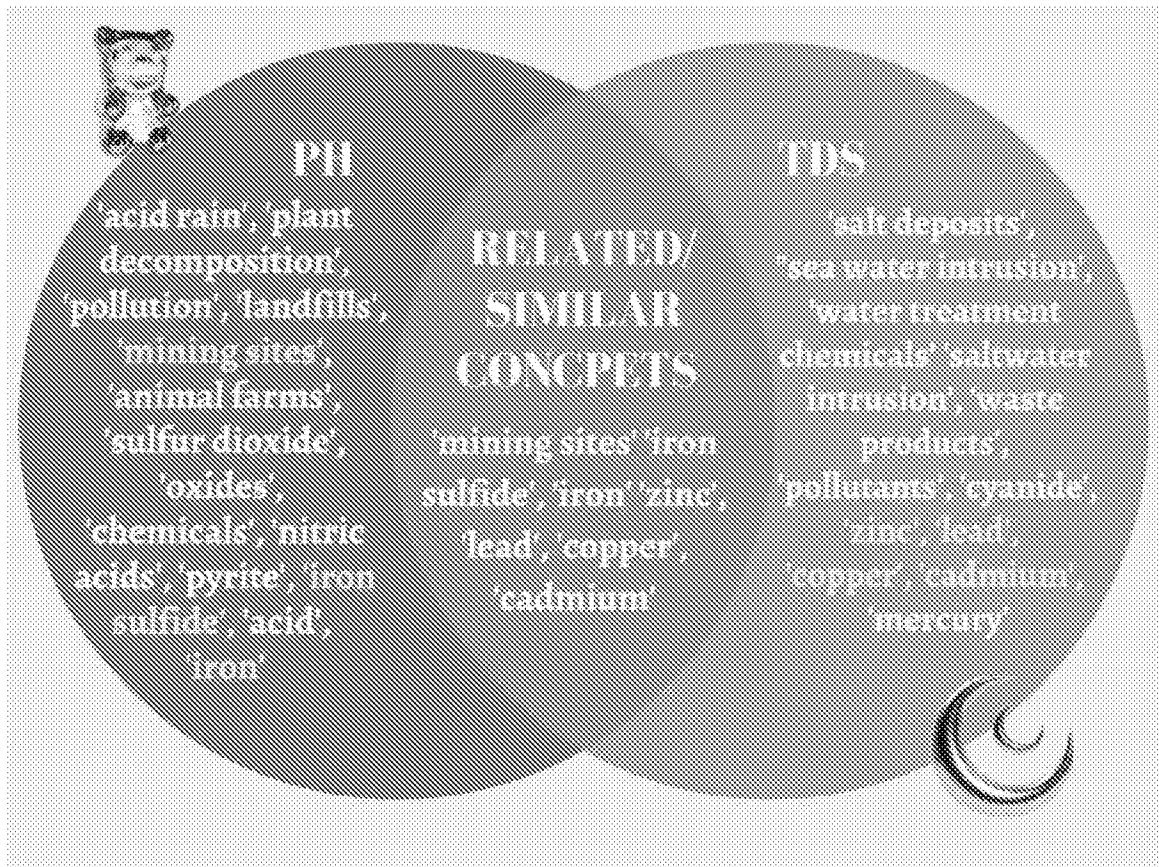
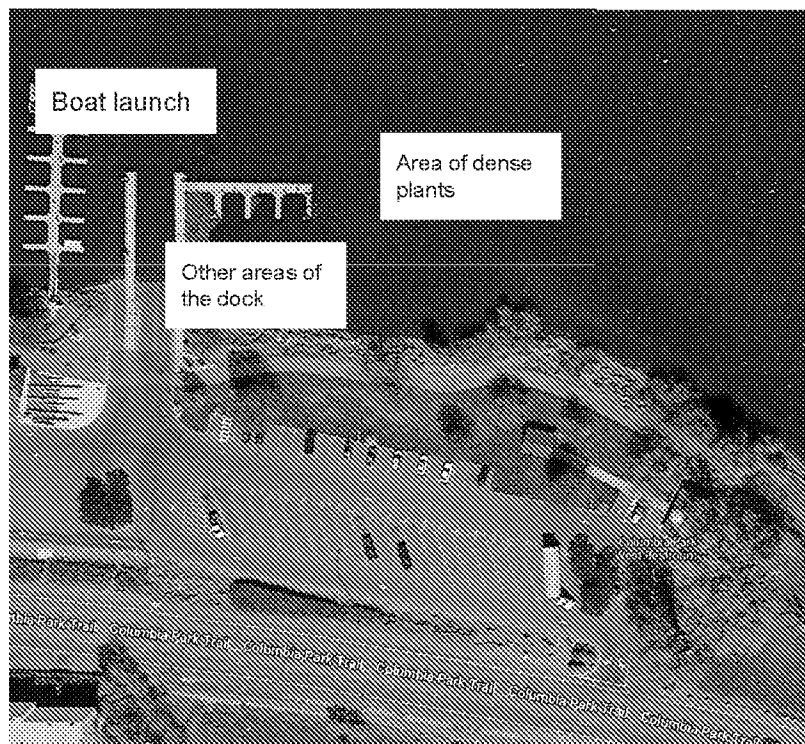


Fig. 14: Illustration of final output based on sensor measurements

	Leslie Groves	Columbia East Pond	Columbia Point Dock
Temperature of water	33 F	35 F	37 F
Ph of water	8.2	10.5	9.4
TDS of water	32	180	29
Location hypothesis	Traces of metal in water due the presence to old, rusting barge	Measurements of waste due to large duck population	Presence of an algae called stargrass near the docks

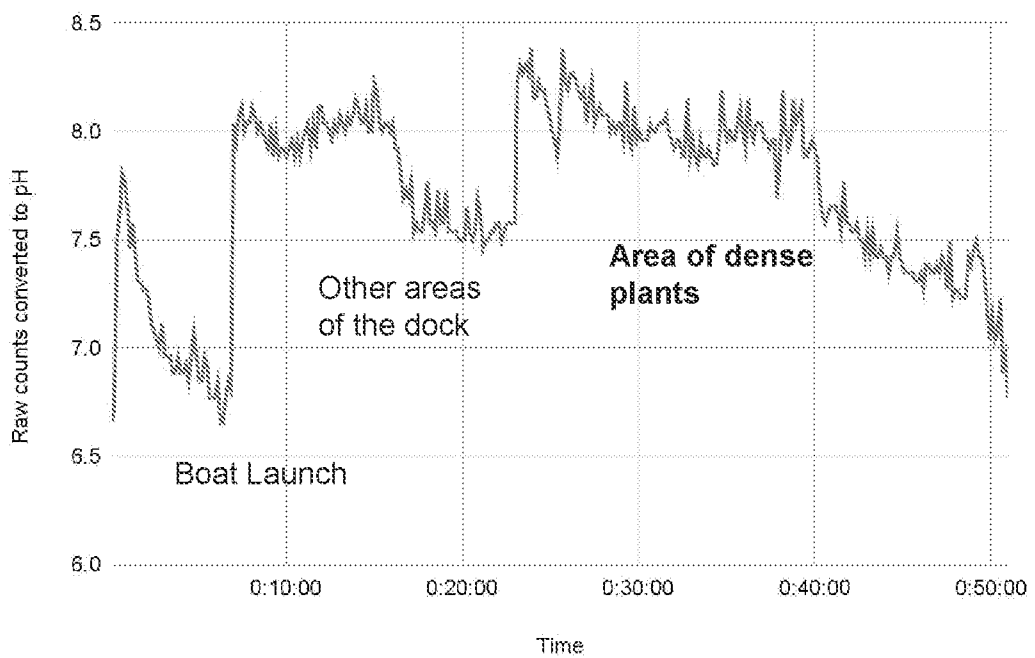
Fig. 15: Hypothesis for stargrass field experiment

Growth of underwater stargrass (invasive species) affects water quality and consequently is impactful for agriculture and fish health



Field experiment: Stargrass data collection in Kennewick, WA

Fig. 16: Data analysis from stargrass field experiment



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/030978

A. CLASSIFICATION OF SUBJECT MATTER		
IPC: <i>G01N 33/18</i> (2024.01); <i>G06F 16/245</i> (2024.01); <i>G06F 16/248</i> (2024.01); <i>B63G 8/00</i> (2024.01); <i>G01N 1/10</i> (2024.01); <i>G06F 17/00</i> (2024.01)		
CPC: <i>G01N 33/18</i> ; <i>G01N 1/10</i> ; <i>B63G 8/00</i> ; <i>G06F 16/245</i> ; <i>G06F 16/248</i> ; <i>G06F 17/00</i>		
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) See Search History Document		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched See Search History Document		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) See Search History Document		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2018/0217029 A1 (WOODS HOLE OCEANOGRAPHIC INSTITUTION) 02 August 2018 (02.08.2018) entire document	1
Y	CN 211925820 U (GUANGZHOU HUANFENG ENERGY TECHNOLOGY CO LTD) 13 November 2020 (13.11.2020) see machine translation	1
A	US 2022/0205966 A1 (CURRENTA GMBH & CO. OHG) 30 June 2022 (30.06.2022) entire document	1
A	US 2020/0346726 A1 (POWERSVISION TECH INC.) 05 November 2020 (05.11.2020) entire document	1
A	CN 115905450 A (SHENZHEN LIANHE INTELLIGENT TECHNOLOGY CO LTD) 04 April 2023 (04.04.2023) see machine translation	1
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“D” document cited by the applicant in the international application</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>		
Date of the actual completion of the international search 09 July 2024 (09.07.2024)		Date of mailing of the international search report 12 July 2024 (12.07.2024)
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300		Authorized officer MATOS TAINA Telephone No. 571-272-4300