RSU Journal of Biology and Applied Sciences

ISSN:

RSUJBAS

About us

Rivers State University Journal of Biology and Applied Sciences (RSUJBAS) publication is a quarterly, open access, international journal for all academic research in science discipline. Microbiology, Botany, Zoology, Environmental Biology, Chemistry, Physics, Mathematics, Computer Science, Biochemistry, Medical Laboratory Sciences and other applied science related areas. RSUJBAS is a platform set for elites to influence, contribute and communicate to the global environment through their various academic researches. We synergistically engage our noble efforts to contribute to the knowledge development, discoveries and innovations in all fields of study. In RSUJBAS we publish research papers on current academic issues with standard scientific reviews. RSUJBAS publishes original research articles, review articles, case studies, short communications, survey report, comparative studies and many more.

Aim and scope

Rivers State University Journal of Biology and Applied Sciences aims to publish high quality papers that communicate fundamental and contemporary discoveries both theoretical and practical. Most importantly RSUJBAS seeks to establish a platform for communicating emerging trends in various discipline such as Microbiology, Botany, Zoology, Environmental Biology, Chemistry, Physics, Mathematics, Computer Science, Biochemistry, Medical Laboratory Sciences and other applied science related areas.

Description:

- Area of concentration: All science academic disciplines
- Frequency of publishing: Quarterly
- Mode of publishing: both online and print publication
- Language of publication: English
- Double Blinded Review Process
- Zero Level Plagiarism Tolerance

Why publish with us

Low Article Processing Charge (ACP) to promote the research work
Easy and Rapid review process
Instant publication upon acceptance
Dedicated editorial and review team for fast review process
RSUJBAS provides hard copies of publication every quarterly

Editorial Board

All Correspondence to
Dr. Sam Wenedu (Editor-in-Chief)
Department of Microbiology, Rivers State University
Editor.jbas@yahoo.com

Or

OLUCHI DICKSON
Publication Manager
Dicksonoluchi87@gmail.com

Consulting Editors

Guideline for Manuscripts

Manuscripts should be typewritten on an A4 sheet having B1.5= line spacing throughout the text. The margins should be 2B54cm (1 inch) in all sides and page number should be consecutively on the bottom of the page. The manuscript should be written in Times New Roman using '12' font size.

For original research paper, the manuscript should be arranged in the following order: Title page, Abstract, Keywords, Introduction, Materials and Methods, Results, Discussion, Acknowledgment, References, Tables with legends, figures with legends and supplementary materials

The title page should contain the title, the name(s) of the author(s), the name(s) and address(es) of the institution(s) where the work was carried out, including a valid e-mail address from the corresponding author along with telephone numbers. The title of the manuscript should be specific and concise but sufficiently informative.

The Abstract should not exceed 250 words and it should contain brief summary of the findings including brief introduction, methodology, results, and conclusions,

The keywords should have a minimum of five and maximum of seven words.

The introduction should provide a clear statement of the problem and indicates aim of the study citing relevant literature to support background statements.

The Materials and Methods should include the methods and methodology of the research.

The Results should be presented in the form of tables or figures. It should be presented with clarity and precision. Statements used to present results should be written in the past tense. Detailed interpretation of data should not be included in the results but should be put into the Discussion section.

The Discussion should interpret the results clearly and concisely, and should integrate the research findings of this and past studies on this topic. Highlight the significant/unique findings of the research under conclusion.

The acknowledgments of people, grants or funds should be brief.

Contents

List of Contributors

Kingsley-Opara, Ngozi
Research Scholar, Department of Computer Science, Ignatius Ajuru University of Education, Rivers State, Nigeria.
Advanced Database Management

Prof. Asagba, Prince Oghenekaro.
Visiting Scholar, Department of Computer Science, University of Port-Harcourt, Rivers State Nigeria.
Email:asagba.prince@uniport.edu.ng

Gabriel B.C., Gabriel M. N, P. O. Asagba
School of Graduate Studies Ignatius Ajuru University Of Education (IAUE), Rumuolumeni, Port Harcourt, Rivers State, Nigeria.
Department Of Computer Science.
gabrielbariyira@gmail.com, meegabz@gmail.com

WAIDOR, Tamaramiebi Keith[1]& ASAGBA, Prince Oghenekaro[2]
Department of Computer Science,
Faculty of Natural and Applied Sciences
Ignatius Ajuru University of Education, Port Harcourt
zalimaxxx@gmail.com

[2]Department of Computer Sciences,
University of Port Harcourt, Port Harcourt, Nigeria
Prince.asagba@uniport.edu.ng

Fiberesima, Alalibo Ralph
Visiting Scholar, Department of Computer Science,
University of Port-Harcourt, Rivers State Nigeria.
fiberesima.a.r@outlook.com;

Asagba, Prince Oghenekaro
Visiting Scholar, Department of Computer Science,
University of Port-Harcourt, Rivers State Nigeria.
asagba.prince@uniport.edu.ng

Kingsley-Opara, Ngozi
Research Scholar, Department of Computer Science,
Ignatius Ajuru University of Education, Rivers State, Nigeria.
Email: ngozikopara@gmail.com

Prof. Asagba, Prince Oghenekaro.
Visiting Scholar, Department of Computer Science,
University of Port-Harcourt, Rivers State Nigeria.

A Model to Compare Web Database and Search Engine

Fiberesima, Alalibo Ralph
Visiting Scholar, Department of Computer Science, University of Port-Harcourt, Rivers State Nigeria.
fiberesima.a.r@outlook.com;

Asagba, Prince Oghenekaro.
Visiting Scholar, Department of Computer Science, University of Port-Harcourt, Rivers State Nigeria.
Email:asagba.prince@uniport.edu.ng

Abstract
Web databases are the predominant and current evolution of the various database architectures and support the ACID and BASE theorems as well as relational algebra and calculus. Search engines are the most recent implementation of the theories behind information retrieval, and in addition to structured data existing in web database, which make up the generality of data available on the Internet, in addition to searching for structured data, are capable of querying unstructured data available in heterogeneous form on either web databases or present in Hypertext Markup Language (HTML) documents, and there exists an intersect between them. The report is based on literature regarding both aspects of Information Systems and is tailored for a Strength, Weakness, Opportunities and Threats analysis of the existing literature on both topics, while laying the theoretical framework for further practical work on the topic. The report uses an exploratory and analytical methodology for assessing web databases and search engines

Keywords: database, information retrieval, SQL, NoSQL, ACID, CAP Theorem, ranking algorithm

Introduction
A web database can be described as a system for storing and displaying information which is retrieved by users of the database from the Internet or world wide web. This is the primary strength of the database and helps it perform any of a number of functions due to its versatility. Such functions include the storage of client details, inventory details and other persistent data. A web database is a type of Web Application and can be grouped as a service platform.

A web database is ideal for situations when the information should be shared, or when it must be accessed from various locations. It is especially beneficial when the system is to be shared between locations or different devices (like tablets, computers and cell phones).
A web database is also described as a database accessible from a local network or the Internet, as opposed to one that is stored locally on an individual computer or its attached storage. Web databases are hosted on websites, made available as software as a service with products accessible via a web browser. These databases may be free or require payment, such as by a monthly subscription as part of the hosting service.

One benefit of a web database in data storage is rather than keep information from a database at one location, a business may choose to have it hosted on the Internet so that all its users can access and update it. Most database services offer web-based consoles, which the end user can use to provision and configure database instances.

The distributed deployment model of the web database has the advantages of improved fault tolerance, as no Single Point of Failure (SPoF) exists, and separate data and application logic allows for at least a theoretical performance boost, but the drawbacks included a more complicated framework, staleness and redundancy of data updates, increased costs due to more components and a lack of best practices for data and application partitioning across computers and locations.

The common availability of the web browser led to the migration of business computer systems to utilization of web pages as the primary mode of information presentation. This is the N-tier client/server model (also known as the Internet computing model) utilized by web databases today.
Though the N-tier client/server model is the dominant and most successful model, as it offers an industry-standard presentation method of HTML, utilizes the same architecture for internal (intranet) and external (Internet) applications while retaining all the benefits of the three-tier client/server model, it however has the disadvantages of inherent lack of security of data on the Internet, the need for larger development project teams with resultant specialization, a higher overall cost, due to additional components and cost of administration.
Aside from the web database having the aforementioned attributes in addition to a standard API for database connectivity based on either open standards or proprietary standards, the advances in database technology led to the need for databases capable of storing the following types of data:

1. Temporal data or current states of real world entities,
2. Spatial data or geographic or computer aided design data composed of points, raster and vector data
3. Multimedia data or image, audio and video data

while retaining the ability to store and retrieve these data in real time, as response is integral to mobile computing (Silberschatz, Korth, & Sudarshan, 2006).

The field of information retrieval and search engines in particular developed in parallel with the field of web databases (Silberschatz, Korth, & Sudarshan, 2011) and early examples are catalogs and document management systems which in turn are composed of HTML pages. Information retrieval systems by design have used a simple model for document retrieval which supposes the internet, correctly, to be composed of unstructured data. Another feature of information retrieval system is the use of keywords for queries which result in domains of knowledge as all results relevant to the keyword are returned to the user.

Search engines have been described as an interactive tool for information retrieval (Liaw and Huang, 2006) from the internet as well as a major tool for searching for information over the internet (Wirth et al, 2007).
The primary features of an information retrieval system are as follows:

1. Indexing; the use of inverted files to map a keyword to the list of documents that contain the keyword
2. Searching; which is the extraction of relevant information from the index
3. Ranking; which makes use of additional parameters such as heuristics to sort the results as closely as possible to match user requirements (Hema Priya &RangaRaj, 2013).

Search engines can be described as an implementation of algorithms for information retrieval through querying of textual and other data, using hyperlink-based techniques. Web databases are the primary hosts for search engines as their structure makes it easier to parse them for relevant results.

An important feature of search engines is the ability to aggregate and rank entries of the returned documents using ranking algorithms. The ability to aggregate entries returned by the search engine must however not be at the cost of space or time needed for such document indexing or query processing (Agrawal et al., 2009) because this would lead to user disinterest either due to excessive volume of results for user parsing leading to a cognitive burden on the user or the engine may take too long to return answers to the search. Furthermore, the sorting heuristics may result in "filter bubbles" (Hoang, Spognardi, Tiezzi, Petrocchi, & Rocco, 2015) which result in the exclusion from the search of documents deemed by the search engine not to fit the users need, an undesired effect for users of the search engine. Sorting heuristics utilized by search engines are the user profile, geo-location of the computer gathered from the IP address, system related data and server side technology such as cookies and web server or database server originating address (Hoang et al, 2015). These features have been adopted for modelling queries based on the personal features they reveal about the user (Zhou, 2015).

Web databases are a very popular topic of research. Their use in multimedia, geographical information systems, and digital libraries demand special requirements with respect to the database schema and instances they are composed of (Ramakrishnan &Gehrke, 2000) as they need schema different from traditional columns and rows. Furthermore, the new data types being stored in web databases rapidly accumulate in their respective domains of use and the challenge of managing these data is the focal point of web database research especially when it concerns information retrieval. The web also has the restriction of usage being open to everyone and not a specialized set of individuals. The common objectives of databases and search engines are the support of information retrieval, however there are points of difference between them in terms of approach to the same concept of data storage and retrieval.
Searches vs queries: Searches are a specialized class of queries that search engines support. They are specified in terms of search terms, with a collection of unstructured text documents comprising the underlying data. Search results may be ordered, based on the closeness of match.

Web database systems support a broad query set, and are composed of rigidly structured data. They have always returned sets of results without any ranking order and this is preserved across different database implementations, even with additional extensions such as PL/SQL and T/SQL because relational queries are precise, with similarity operators the only way of adding similar values and returning multiple rows. Apart from these, only two answers are possible for any query, "yes" and "no" (Silberschatz, Korth, & Sudarshan, 2011). have been tailored to be read-only with respect to their workload and the concept of transactions is alien to them. Search engine systems work using periodic document addition to an existing index of the document collection, and these index structures are rebuilt or updated at intervals based on the algorithm. The index structures affect how documents can be retrieved or not because the freshness or staleness of such indexes has a direct outcome on the relevance of the search. They support boolean queries constructed from operators as well as ranked queries which are based on the following theorems

Ranking Using TF-IDF: here the relevance of a particular document to the search term is evaluated.

Similarity-Based Retrieval: here the retrieval of documents for the user is based on the similarity to the search term. This method takes cognizance of all documents as points and

vectors in an n-dimensional space called the vector space model (Silberschatz, Korth, & Sudarshan, 2011). The resultant document set based on the idea of relevance feedback, with the search engine suggesting likely documents to the user.
In contrast, web database systems support update-intensive transaction processes in a transparent manner using the ACID or BASE theoretical frameworks for processing workloads.
Relevance using hyperlinks: this is subjective to the accuracy of information retrieved from the host of a website, but the general idea is the number of times a page is accessed with respect to a search query many times a page is a measure of the site's popularity. Search engines retrieve documents by a process called crawling, and through analysis find links between the pages for relevance to the query. Crawling is the browsing of the world wide web in a manner that is both analytical and methodical. It is a synonym for a web spider.
The transfer of prestige principle however is a simpler means of similarity as a link from a popular page to another page is treated as conferring more prestige than a link from a not-so-popular page. This theorem has given rise to the Page Rank that is based on the random walk model. However, the Page Rank algorithm has the drawback of not take query keywords into account in the assignation of popularity.

A main disadvantage of the Page Rank algorithm is search engine spamming, the practice of creating web pages designed to get a high relevance rank for some queries, in spite of the lack of popularity or relevance of those sites.

Materials and methods

The materials used in the study were software comprising of web databases, specifically databases provided by hosting services as part of their paid and free tiers were examined. The present study was carried out to know the theory and practice regarding the use of web databases and search engines. The analytical method of deconstructing the documentation and the exploratory method of Literature Review (LR) is used as a tool to construct a base of further research questions in web databases and search engines which are formulated from the identification of factors revealed via analysis of the literature in the aforementioned topics

Results

Based on the literature a model of comparison can be made between web databases and search engines. The model which seeks to compare them is based on criteria as seen in the table below:

	Web database	Search Engine
Persistent collection	Yes	No
Indexing	Structured	Unstructured
Data Search	Yes	Yes
Query retrieval	Yes	Yes
Result format	Limited	Unlimited
Scope	Restricted to query	Full text search
Natural Language Search Query	No	Yes
User proficiency	Skilled	Ordinary
Specialised Language	SQL	Natural language

Table 1: Comparison between web database and search engine

Discussion

The world wide web can therefore be said to be a meeting ground between web databases and information retrieval systems.
As earlier stated, search engines have the ability to combine other search engines as part of their information retrieval algorithm, though this does not guarantee an improvement in the relevance of the results generated by such systems.
Search engines are a result of research into information retrieval systems, especially of the web databases that make up the internet. These hold a massive volume of data, and often such data is composed of not only the instance and schema, but also unstructured data. Tools such as spiders which analyze links from the different sites on the internet and analyze the content of the links for relevance have been part of information retrieval since the inception of information retrieval systems and are the basis of web portals being visible to users of the internet. An important aspect of search engines is their ability to extricate meta-data from the internet search results otherwise known as documents which includes the URL, page title among other information.

The categorization of pages which contain information relevant to a query is also a function of search engines, and commonality of results returned by different search engines can also be a means for assessing the relevance of such engines with respect to a particular search query.

Web databases traditionally support all the relational and NoSQL architecture including Atomicity, Consistency, Isolation, Durability, Relational algebra, Relational calculus (domain and tuple calculus), Basically Availability, Soft state and Eventual consistency as a result of gradual propagation of data across all nodes in the database implementation (Sadalage& Fowler, 2013). However search engines support boolean or ranked queries and use specialized information retrieval algorithms tailored to suit that particular implementation
Web database implementations include MySQL and MongoDB which are based on the RDBMS and NoSQL CP implementation respectively. CP is taken from two letters of the CAP theorem, maintaining consistency, while compromising on availability in resolution of network partitions. MongoDB uses replica sets which can only have one primary node that receives writes and propagates them to the secondary nodes.

The MySQL InnoDB storage engine supports and guarantees ACID-compliant features.
Search engine implementations include Google, Bing and DuckDuckGo and each has its algorithm for page rank and document retrieval respectively, which may be proprietary or open source.

Web databases and search engines are ubiquitous on the web, Web databases like MySQL and MongoDB support their respective mechanisms for storage and transaction fidelity. Search engines, like Google and Yahoo support similarity and boolean document retrieval mechanisms based on the particular implementation. In this paper, we started from previous work in the area and continued to investigate the reason behind web databases and search engines.
The usefulness of knowledge of the intersection between web databases and search engines has also been discussed and the importance and drawbacks of domain-specific queries and search personalization methods.
Further study suggestions is the study of different techniques and practical implementation of web database and search engine theories and implementations

Acknowledgments

I acknowledge the contributions of the reviewers of the technical and documentary work that went into the production of this journal.

References

Agrawal, S., Chakrabarti, K., Chaudhuri, S., Ganti, V., Konig, A. C & Xin, D. (2009). Exploiting web search engines to search structured databases. *In Proceedings of the 18th international conference on World wide web (WWW '09)*. Association for Computing Machinery, New York, NY, USA, 501–510. DOI:https://doi.org/10.1145/1526709.1526777

Hema Priya, P., &RangaRaj, R. (2013). An Improved Search Engine by Semantic Web using Ontology, *International Journal of Science and Research (IJSR),2* (12), pp. 403 - 408. Retrieved from https://www.ijsr.net/search_index_results_paperid.php?id=02013583

Hoang, V. T., Spognardi, A. Tiezzi, F., Petrocchi, M..&Rocco, D. (2015). Domain-specific queries and Web search personalization: some investigations. *Electronic Proceedings in Theoretical Computer Science.188*. DOI:10.4204/EPTCS.188.6.

Liaw, S., & Huang, H.. (2006). Information retrieval from the World Wide Web: a user-focused approach based on individual experience with search engines. *Computers in Human Behaviour, 22* (3),501-517.

Oppel, A. (2010). Databases Demystified, 2nd ed. McGraw-Hill Professional, Boston, MA

Ramakrishnan, R., &Gehrke, J. (2000). Database Management Systems, 2nd ed. McGraw-Hill Higher Education, Boston, MA

Ramakrishnan, R., &Gehrke, J. (2002). Database Management Systems, 3rd ed. McGraw-Hill Higher Education, Boston, MA

Resources for developers, by developers. (2020, May 20). Retrieved from https://developer.mozilla.org/en-US/

Rob, P., & Coronel, C. (2007). Database Systems: Design, Implementation, and Management, 8th ed. Course Technology, Boston, MA

Sadalage, P. J., & Fowler, M. (2013). NoSQL Distilled. Addison-Wesley Professional, Boston, MA

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011). *Database system concepts*. New York: McGraw-Hill.

Stewart, J. M., Chapple, M., & Gibson, D. (2015). CISSP (ISC)2 Certified Information Systems Security Professional Official Study Guide, 7th ed. John Wiley & Sons, Indianapolis, IN

Tiwari, S. (2011). Professional NoSQL. John Wiley & Sons, Indianapolis, IN

Williams, H. E., & Lane, D. (2004). Web Database Applications with PHP and MySQL, 2nd ed. Sebastopol, CA

Wirth, W., Bocking, T., Karnowski, V., & Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication, 12* (2), 778-800.