International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Location Based Ensemble Classified Site Using Search Based Annotation Technique

[1]PL. Muthukumar, [2]Dr. R. Karthiyayini

[1]Student, Master of Computer Applications. University College of Engineering, Anna University, BIT Campus, Trruchirappalli.
[2]Assistant Professor, Sr. Guide, Master of Computer Applications. University College of Engineering, Anna University, BIT Campus, Trruchirappalli.

ABSTRACT-

Classified Management System is a real-time Python-based project. Technology always evolves over time. This website is made using the latest technologies such as python and django platforms to make shopping fast, convenient and hassle-free for everyone anywhere with a working internet connection. Anyone who needs a product from the company can not only visit the ion store, but also shop online. On this site, the user can consult the product without registering an account or entering data, so that he can return to access it at his convenience and continue where he left off. A user account that provides all the data required to place an order, which will then be delivered to you after successful payment. Additionally, the site offers location-based searches, keyword annotation, and quick content searches, and shopping by category has many filters that make it easier to distill categorical data. The proposed work is to implement the use of python and django and use different latest libraries, flask framework to improve the user's graphical view experience.

Keywords: classified, e-commerce, keyword, annotation, python, location.

INTRODUCTION

Today, many e-commerce sites personalize their content, including Netflix (movie recommendations), Amazon (product recommendations), and Yelp (trade reviews). In many cases, customization benefits users: For example, when a user searches for an ambiguous query like "router," Amazon might suggest woodworking tools instead of networking equipment. However, personalization on e- commerce sites can also harm users by manipulating displayed products (price targeting) or personalizing product prices (price discrimination). Unfortunately, today, we lack the necessary tools and techniques to detect such behavior.

Recently, researchers and internet users found evidence of personalization of e-commerce sites. On such sites, the benefits of personalization to users are less obvious; e- commerce sites have an economic incentive to use personalization to entice users to spend more. For example, travel site Orbitz was found to personalize hotel search results. Unbeknownst to users, Orbitz "directs" Mac OS X users to more expensive hotels in specific locations by getting them to rank higher in search results. After a month, Orbitz stopped using this personalization algorithm. Concerned about filter bubbles, our previous work set out to explore what triggers personalization in Google Search.

We've found that Google infers a user's geographic location from their IP address, and location-based personalization affects search results more than any other single feature.

However, while these initial findings are interesting, many questions remain, such as: Does location-based personalization affect all query types (eg, politics vs news) equally? At what distance do users start seeing changes in search results by location? Answering these questions is critical because a user's geographic location can be used as an indicator of other demographic characteristics, such as ethnicity, income level, education level, and political affiliation.

PREVIOUS WORK

Most existing reclassification methods use a tool called pseudo-relevant feedback (PRF), where a subset of the top- ranked images are assumed to be relevant and subsequently used to build a reclassification model. This contrasts with relevance feedback, where users explicitly provide feedback by marking top results as positive or negative. In classification-based PRF methods, the best-classified images are treated as pseudo-positives, and the poorly-classified images are treated as pseudo-negatives to train the classifier, and then reclassify. Hsu et alv. This pseudo-positive and pseudo-negative image approach was also adopted to develop a clustering-based reclassification algorithm.

PROPOSED WORK

The application focuses on performing XML searches using fuzzy search rather than look-ahead. The system allows users to identify XML documents whose content should be searched. Users can create XML files from SQL Server databases at runtime. The user must pass parameters to the SQL server,

which authenticates the user and provides access to the source tables from which the XML file was created. Users search using type-ahead search, which suggests words as the user types in Google. However, this is a limited option and may not meet everyone's needs. The system suggests fetching the data using Fuzzy, which performs character insertion/update/deletion to generate an extensive set of words and provide corrections in the case of misspelled words. Most lookups examine the values and determine their location in the XML file. However, the system suggests a search value or tag or tag attribute or tag attribute value. During the search process, the item being searched is assumed to be unique, while at other times it is ambiguous. There is no ambiguity in direct search, user will get details of search data. If there is ambiguity, the search- through pattern is used.

LITERATURE REVIEW

1. **Wahid et. al., (2021):** Adding semantics to existing web content is one of the most important steps to take to fully realize the Semantic Web, and remains a significant challenge. Much of the research in this area has focused on extracting information from various types of data sources, including news, emails, Wikipedia, Internet taxonomies, and more. The reported techniques are well investigated.
2. **Koltcov et. al., (2020):** Described topic modeling in the field of machine learning as a statistical model for discovering abstract topics and is one of the text mining techniques used to discover the hidden semantic structure of text.
3. **Abayomi-Alli et. al., (2022):** Analyzed Symmetric key algorithms check for differences in encryption time. Analysis showed that the encryption time was independent of the data type and data density of the files. So encryption only depends on the number of bytes in the file. In addition, the proportional relationship between encryption time and data size is explored. When the data size increases, the encryption time also increases, and vice versa. Increasing the key size can improve encryption time, but it decreases with increasing key size for all block cipher algorithms. It can be seen from the simulation results that the encryption time does not depend on the data type and file density. Using a key algorithm increases the time required to encrypt data.
4. **Li et. al., (2022):** Enlightened fuzzy logic is useful when trying to mechanize or formalize human capabilities. First, the ability to converse, reason, and make rational decisions in an environment where information is imprecise, uncertain, conflicting, and incomplete. Second, the ability to perform a variety of physical and mental tasks without any psychometric measurements or calculations. There are some criteria for selecting the best insurance company among a group of companies. Criteria have some independent weighting values. Evaluate the best insurance company alternatives based on a set of weighted criteria. The insurance company's alternatives were considered for final implementation, which was the best option evaluated against all other criteria.

ALGORITHMS USED

A. Geo Location Based Search (K-NN Algorithm):

Step 1 − Load Users geo location (Latitude and Longitude).

Step 2 − Next, to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 − For each point in the test data do the following-

3.1 − Calculate the distance between two locations using Euclidean, distance.
3.2 − Now, based on the distance value, sort them in ascending order.− Next, choose the top K rows from the sorted array.
3.3 − Now, assign a class to the test point based on most frequent class of these rows.

Step 4 − End

B. Click Prediction Re-ranking using Decision Tree

Step-1: Begin the tree with the root node, says S, which contains the complete category list.

Step-2: Find the best attribute in the dataset using Click Through Points (CTR)

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

EXPERIMENTAL RESULTS

A). Fuzzy Search: The fuzzy keyword search technology mentioned above is implemented in Python. Python is chosen as the programming language, and other open source APIs (Application Programming Interfaces) support other functions. Pycharm is used as a development IDE (Integrated Development Environment), and libraries for other technologies are added to pycharm as external packages. Built on the python platform, Pycharm integrates proprietary and open source solutions into the development environment. Py Plot is an open source framework for the Python programming

language. It is an open source library available that allows users to easily generate graphs and charts. It is especially effective when the user needs to regenerate a graph that changes frequently.

TABLE I. EFFICIENCY IN SEARCHING FUZZY KEYWORD

Training Words	Gram Based	Wildcard Based	Symbol Based
100	98.33	97.66	99.33
500	97.33	96.33	98.5
1000	95.33	96.33	97.63
2500	95	95.89	97.33

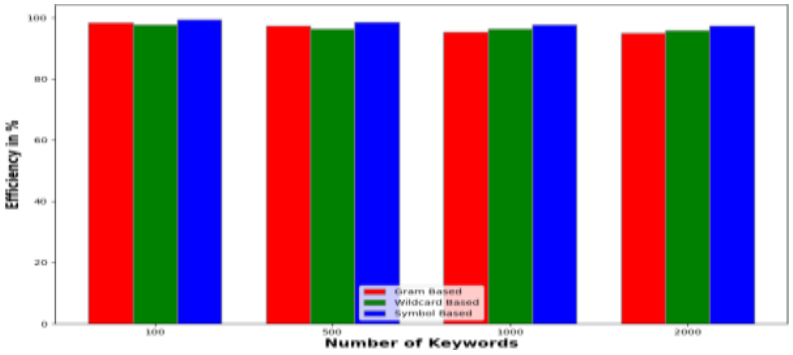

Fig. 1. Performance chart for Fuzzy Keyword Search

B) Geo Location Based Search: This paper first investigates whether the aggregated frequency distribution of functional location codes is the same as location tracking. The frequency distributions of location tracking and time usage data shown in Table 2 are more similar for network users. This is largely due to the excess of other categories of web users based on location tracking, which account for nearly half of web user registrations. Other functional location categories are also overrepresented in the data, but far below the web data. For web users, the percentages of location- tracked households and public transit were about the same.

TABLE II. EFFICIENCY IN GEO LOCATION BASED SEARCH

Range	Time Use (ms)	Location Tracking (%)
Within 1 km	3.7	98.66
Within 5 km	5.2	96.33
Within 15 km	7.8	94.76

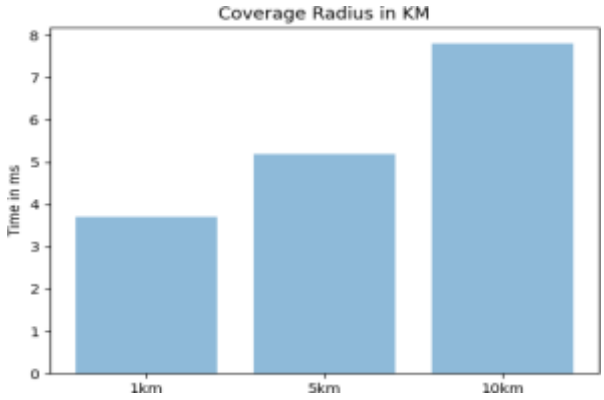

Fig. 2. Time difference chart for Geo Location Based Search

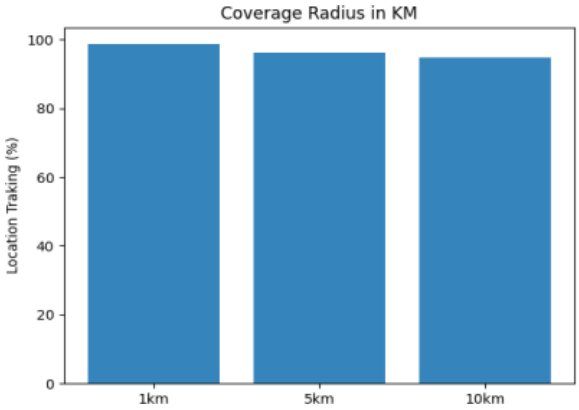

Fig. 3. Location Tracking chart for Geo Location Based Search

CONCLUSION

In this work, the problem of efficient XML and fuzzy keyword search is studied, including the identification of user search intent and the classification of results with keyword ambiguity. I use statistics to infer users' search intent and rank query results. In particular, XML TF and XML DF are defined, based on which we design formulas to calculate the confidence of each candidate node type for lookup/search by nodes, and further propose a novel node classification scheme. Captures the hierarchy of XML data. These sets are then used to perform fuzzy keyword searches on encrypted data using the scheme proposed in this paper.This solution is proven to improve security while preserving privacy. It also improves the overall performance by increasing the operation speed, so as to achieve the purpose of fuzzy keyword search.

Reference

[1]. Saxena, Krati, Sagar Sunkle, and Vinay Kulkarni. "Hybrid Search based Enhanced Named Entity Annotation Tool." 15th Innovations in Software Engineering Conference. 2022.

[2]. Kundaje, Anshul, and Wouter Meuleman. "Automated sequence- based annotation and interpretation of the human genome." Nature Genetics (2022): 1-2.

[3]. Mohamed, Ensaf Hussein, and Eyad Mohamed Shokry. "QSST: A Quranic Semantic Search Tool based on word embedding." Journal of King Saud University-Computer and Information Sciences (2020).

[4]. Tamer, Nazif Can, and Murat Saraçlar. "Improving keyword search performance in sign language with hand shape features." European Conference on Computer Vision. Springer, Cham, 2020.

[5]. Kummervold, Per E., et al. "Categorizing vaccine confidence with a transformer-based machine learning model: analysis of nuances of vaccine sentiment in Twitter discourse." JMIR medical informatics 9.10 (2021): e29584.

[6]. Taub-Tabib, Hillel, et al. "Interactive extractive search over biomedical corpora." arXiv preprint arXiv:2006.04148 (2020).

[7]. Anand, Lakshay, and Carlos M. Rodriguez Lopez. "chromoMap: an R package for interactive visualization and annotation of chromosomes." Biorxiv (2020): 605600.

[8]. Ishida, Youichi, Toshiyuki Shimizu, and Masatoshi Yoshikawa. "An analysis and comparison of keyword recommendation methods for scientific data." International Journal on Digital Libraries 21.3 (2020): 307-327.

[9]. Bandari, Sumalatha, and Vishnu Vardhan Bulusu. "Survey on ontology-based sentiment analysis of customer reviews for products and services." Data Engineering and Communication Technology. Springer, Singapore, 2020. 91-101.

[10]. Shi, Yuxuan, Gong Cheng, and Evgeny Kharlamov. "Keyword search over knowledge graphs via static and dynamic hub labelings." Proceedings of The Web Conference 2020. 2020.

[11]. Gligorijević, Vladimir, et al. "Structure-based protein function prediction using graph convolutional networks." Nature communications 12.1 (2021): 1-14.

[12]. Köhler, Niklas D., et al. "Deep learning does not outperform classical machine learning for cell-type annotation." bioRxiv (2021): 653907.

[13]. Li, Xiner, et al. "Keyword Search Based on Unsupervised Pre- Trained Acoustic Models." International Journal of Asian Language Processing (2022): 2250005.

[14]. Gholamrezaie, Faeze, Melika Bahman-Abadi, and M. B. Ghaznavi- Ghoushchi. "AI Annotated Recommendations in an Efficient Visual Learning

Environment with Emphasis on YouTube (AI-EVL)." arXiv preprint arXiv:2203.11157 (2022).

[15]. Kumar, Kopuri Naveen, and Aarti. "A literature survey on bag of visual words model for classification of histopathological images with deep learning." AIP Conference Proceedings. Vol. 2418. No. 1. AIP Publishing LLC, 2022.

[16]. Markarian, Nicholas, et al. "Validation of Machine Learning–Based Automated Surgical Instrument Annotation Using Publicly Available Intraoperative Video." Operative Neurosurgery (2022): 10-1227.

[17]. Sharma, Vaishali, Khyati Ahlawat, and Shorya Sharma. "Practical Approaches: User Stories Domain Prediction and Keyword Extraction for Insights." Computational Intelligence in Machine Learning. Springer, Singapore, 2022. 189-203.

[18]. [18] Fei, Xue. "An LDA based model for semantic annotation of Web English educational resources." Journal of Intelligent & Fuzzy Systems 40.2 (2021): 3445-3454.

[19]. Zhao, Zeyu, and Wei-Qiang Zhang. "End-to-end keyword search system based on attention mechanism and energy scorer for low resource languages." Neural Networks 139 (2021): 326-334.