

# Ontology based Personalized Mobile Search Engine

K. Anita Davamani, S. Amudha

Received: 12 December 2016 • Revised: 15 January 2017 • Accepted: 14 February 2017

**Abstract:** Web search is a frequent activity on Internet connected devices, yet web search is still a nuisance when user is using a mobile device, due to their default keypad and small screen, and that search results could be mostly irrelevant for the user needs given its mobile context. User needs an efficient way to introduce query terms and receive more precise information. In this paper, we propose a new web search personalization approach that captures the user's interests and preferences in the form of concepts by mining search results and their click through. Onto-Search is based on a client server model. Server performs heavy tasks such as training, re ranking, To preserve privacy only the feature vectors is passed to the server. Location information is also taken into consideration. In addition gps locations help in reinforcing search results. Finally, based on the derived ontology SVM is used for re-ranking of future search results.

**Keywords:** Personalization, Content Ontology, Location Ontology, Re Ranking Search Results, Mobile Search Engine, Click through Data.

## INTRODUCTION

In today's era internet is widely used. Using internet through mobile has become common nowadays. But there arises some issues like small keypad, limited display screen. To be efficient mobile search engine must be able to profile user's interests and return highly relevant results by submitting shorter queries. Leung et al developed a system that personalizes search results based on concept preferences and showed that it is more effective than page preferences [2]. Several personalization techniques have been proposed to model users' content preferences via analysis of users' clicking and browsing behaviors [3],[4],[5],[6]. In this paper we personalize search using content and location ontology, user preferences. For example if the user provides a query "hotel", a set of results will be returned to the client. Based on the click through of the user Onto-Search understands that the user's content preference is "reservation" and location preference is "India". Further GPS locations of the user are also tracked. If the user is at the location "Chennai" then it will be effectively used in personalization. For Personalization of future queries the user preference "reservation", "Chennai" will be taken into account. Thus for future queries Onto-Search will favour results concerning hotels in Chennai. GPS locations can be obtained by mobile devices. To provide effective personalization Onto-Search is capable of combining users location preferences and GPS locations.

The Onto-Search client is responsible for receiving the user's requests, submitting the requests to the Onto-Search server, displaying the returned results, and collecting his/her click through in order to derive his/her personal preferences. The Onto-Search server, on the other hand, is responsible for handling tasks such as training and re ranking of search results before they are returned to the client. The profiles of users are stored on the Onto-Search clients, thus preserving privacy to the users. Onto-Search client performs tasks such as submitting the query to the Onto-Search server, providing the results obtained from the server, collecting the click through of the user that can be used in personalization. The Onto-Search performs computationally heavy tasks such as training and re ranking. Onto-Search client is a Google Android platform and Onto-Search server is a PC server. Only the feature vectors obtained from the users click through are passed to the server thus preserving the privacy of the user. These vectors are used for personalization of future queries.

---

K. Anita Davamani, Assistant Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E-mail: anitadavamani@gmail.com

S. Amudha, Assistant Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E-mail: amudha17s@gmail.com

## RELATED WORK

Most search engines return roughly the same results to all users which may be relevant or irrelevant. However, different users may have different information needs even for the same query. For example, a user who is looking for a laptop may issue a query “apple” to find products from Apple Computer, while a housewife may use the same query “apple” to find apple recipes. The personalization of search results of the search engines were based on users click through [7]. There are some proposals which include an ontological user profile and even define a context model to determine user’s interests or re-rank search results according to the profile created while the user is surfing the web [8] [9], but essentially they consider desktop users only. Focusing on the mobile environment, there is previous research related to web search. Some proposals already uses ontology to represent user profile, others include some context elements but usually limited to the geographic location. Some others approaches use ontology to create clusters of results.

There are other works which deal with word recommendation and auto completion[10] on mobile devices, but they rarely include the ontology knowledge. Also, there is a newer propose for a standardized mobile ontology [11] but it is focused on mobile services instead the web search, so it does not include elements relevant to obtain better search results. Rank Net, scalable implementation of neural networks, is then employed to learn the user behavior model from the cleaned click through data. Earlier personalization techniques were based solely on the computational behavior of the user to model his interests regardless of his surrounding environment. The main limitation of such approaches is that they do not take into account the dynamicity of the user interests.

Gan et. al suggested that search queries can be classified into two types, content (i.e.non-geo) and location (i.e.geo).A classifier was built to classify geo and non-geo queries, and the properties of geo queries were studied in detail. It was found that a significant number of queries were location queries focusing on location information. Hence, a number of location-based search systems designed for geo queries have been proposed. These include Yokoji et al search engine who proposed a location-based search system for web documents. A parser was employed to extract location information from web documents, which was converted into latitude longitude pairs or polygons .When a user enters a query together with the location information specified in a latitude longitude pair, the system creates a search circle within the specified latitude-longitude pair and retrieves documents containing location information within the search circle.

## SYSTEM ARCHITECTURE

An ontology-based, multi-facet (OMF) user profiling strategy to capture both of the users' content and location preferences (i.e., .multi-facets.) is used for building a personalized search engine for mobile users. Figure 1 shows the general process of our approach, which consists of two major activities: 1) Re-ranking and 2) Profile Updating. The search results are obtained from the backend search engines (e.g., Google, MSN Search, and Yahoo).

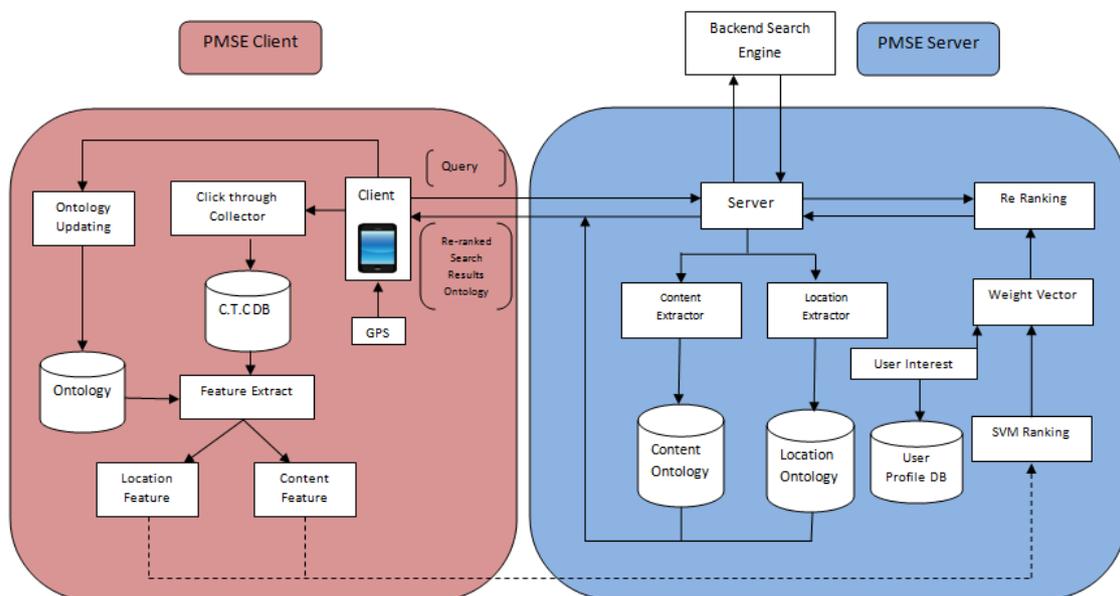


Fig.1: System Architecture

When a user enters a query, The results obtained from the search are combined and re-ranked according to the user's profile trained from the user's previous search activities. After the search results are obtained from the backend search engines, the content and location concepts (i.e. important terms and phrases) and their relationships from the search results and stored, respectively, as content ontology and location ontology.

After the search results are obtained from the backend search engines, the content and location concepts (i.e. important terms and phrases) and their relationships from the search results are stored, under content ontology and location ontology. When the user clicks on a search result, the results, content and location concepts are stored in a database. RSVM training is done to using the feature vectors to obtain a content weight vector and a location weight vector for re-ranking. The search results for the user GPS locations are used along with location preferences for effective personalization. Weight vector is incremented for the particular GPS locations and is updated in user profile. When the user issues the same query again the personalized results according to the user content and location preferences are displayed to the user. Thus the user can get an effective result which helps him in making quick decision making without any confusion.

### **Concept Extraction**

Our personalization approach is based on concepts to profile the interests and preferences of a user. Therefore, an issue we have to address is how to extract and represent concepts from search results of the user. We propose in this paper an ontology-based, multi-facet (OMF) profiling method in which concepts can be further classified into different types, such as content concepts, location concepts. Here we define two concepts namely, content concepts and location concepts. A content concept, like a keyword or key-phrase in a Web page, is related to the content of the page, whereas a location concept provides information about the places related to the page.

#### **Content Ontology**

If a keyword/phrase exists frequently in the web-snippets arising from the query  $q$ , it can be considered as an important concept related to the query. Thus, our content concept extraction method first extracts all the keywords and phrases from the web-snippets arising from  $q$ . The possible concept space determined for the query "hotel," while the click through data determine the user preferences on the concept space. In general, the ontology covers more than what the user actually wants. The concept space for the query "hotel" consists of "map," "reservation," "room rate,"..., etc. If the user is indeed interested in information about hotel rates and clicks on pages containing "room rate" and "special discount rate" concepts, the captured click through favors the two clicked concepts. Feature vectors containing the concepts "room rate" and "special discount rate" as positive preferences will be created corresponding to the query "hotel." When the query is issued again later, these feature vectors will be transmitted to the Onto-Search server and transformed into a content weight vector to rank the search results according to the user's content preferences.

#### **Location Ontology**

In the location ontology, we organize all the cities as children under their provinces, all the provinces as children under their regions, and all the regions as children under their countries [1].The location ontology extraction method first extracts all of the keywords and key-phrases from the documents returned for  $q$ . If a keyword or key-phrase in a retrieved document  $d$  matches a location name in our predefined location ontology, it will be treated as a location concept of  $d$ . For example, given the concept "Tamil Nadu" from document  $d$ , we would match it against our location ontology. If a concept matches several nodes in the location ontology, all matched locations will be associated with the document.

#### **Click through Collection**

When the user issues a query a set of results are returned to the client. When the user clicks on a search result the click through is captured and the clicked concept is favored. The content ontology together with the click through serve as the user profile in the personalization process. They will then be transformed into a linear feature vector to rank the search results according to the user's preferences. These search preferences, along with a set of feature vectors are submitted along with future queries to the server for search results re-ranking.

Spy NB learns user behavior models from preferences extracted from click through data. Spy NB assumes that users would only click on documents that are of interest to them. Thus, it is reasonable to treat the clicked documents as positive samples. However, un-clicked documents are treated as unlabeled samples because they could be either relevant or irrelevant to the user.

Based on this interpretation of click through, the problem becomes how to predict from the unlabeled set reliable negative documents which are irrelevant to the user.

### RSVM

Ranking SVM is an application of SVM to solve certain ranking problems. Its purpose is to improve the performance of the internet search engine. Using click through data as the input, RSVM aims at finding a linear ranking function, which holds for as many document preference pairs as possible. It maps the similarities between queries and the clicked pages onto certain space. It calculates the weights between any two of the vectors obtained. Re-ranks the search results based on the weights. If a user has visited the GPS location, the weight of the location concept is incremented. Hence, we assume that the location that the user has visited a long time ago is less important than the location that the user has recently visited. Thus GPS plays a important role in location information.

### EXPERIMENTAL EVALUATION

Finding the required information quickly and easily on the web remains a major challenge. Most search engines compute a numeric score on how well each search results matches the query, and rank the results according to this value. The top ranking results are then shown to the user.

The users are asked to perform relevance judgment on the top 100 results for each query by filing in a score for each search result to reflect the relevance of the search results to the query. The score indicates three levels of relevancy (Good, Fair and Poor). Documents rated as Good are considered relevant (positive samples), while those rated as Poor are considered irrelevant (negative samples) based on the user's needs. The documents grouped as Fair are treated as unlabeled. Documents rated as Good (relevant documents) are used to compute the average relevant rank improvements (i.e., the difference between the average ranks of the relevant documents in the search results before and after personalization) and top N precisions, the two primary metrics for our evaluation.

Considering our proposed system, we use two straightforward evaluation measures, namely, precision and recall. Different queries may induce from the search results different concept spaces which are different in both sizes and diversities. Additionally, different users may have different interests and preferences on the search results. The diversity of content and location information of a query can be obtained from the extracted content and location concepts, while the interest/preference of a user can be found by his click through behavior. In information retrieval contexts, precision is the fraction of retrieved documents that are relevant to the search. Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

Precision and recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced by a web search engine for a query) and a set of relevant documents (e.g. the list of all documents on the internet that are relevant for a certain topic).

### CONCLUSION

In this paper, we proposed an Ontology-Based Personalization framework for automatically extracting and learning a user's content and location preferences based on the user's clickthrough. Here we develop different methods for extracting content and location concepts, which are maintained along with their relationships in the content and location ontologies. As for the future work, we plan to study the effectiveness of other kinds of concepts such as people names and time for personalization.

### REFERENCES

- [1] Das, J., Das, M. P., & Velusamy, P. (2013). Sesbania grandiflora leaf extract mediated green synthesis of antibacterial silver nanoparticles against selected human pathogens. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 104, 265-270.
- [2] Umanath, K.P.S.S.K., Palanikumar, K., & Selvamani, S. T. (2013). Analysis of dry sliding wear behaviour of Al6061/SiC/Al2O3 hybrid metal matrix composites. *Composites Part B: Engineering*, 53, 159-168.
- [3] Udayakumar, R., Khanaa, V., Saravanan, T., & Saritha, G. (1786). Cross layer optimization for wireless network (WIMAX). *Middle-East Journal of Scientific Research*, 16(12), 1786-1789.
- [4] Kumaravel, A., & Rangarajan, K. (2013). Algorithm for automaton specification for exploring dynamic labyrinths. *Indian Journal of Science and Technology*, 6(5S), 4554-4559.

- [5] Pieger, S., Salman, A., & Bidra, A.S. (2014). Clinical outcomes of lithium disilicate single crowns and partial fixed dental prostheses: a systematic review. *The Journal of prosthetic dentistry*, 112(1), 22-30.
- [6] Vijayaraghavan, K., Nalini, S.K., Prakash, N.U., & Madhankumar, D. (2012). One step green synthesis of silver nano/microparticles using extracts of *Trachyspermum ammi* and *Papaver somniferum*. *Colloids and Surfaces B: Biointerfaces*, 94, 114-117.
- [7] Khanaa, V., Mohanta, K., & Satheesh, B. (2013). Comparative study of uwb communications over fiber using direct and external modulations. *Indian Journal of Science and Technology*, 6(6), 4845-4847.
- [8] Alageswari, P., and Nandhakumar, S.K. (2016). Design of SM Controller Technique for Photo Voltaic System with DC-DC Converter. *International Journal of System Design and Information Processing*, 3(1), 1-5.
- [9] Dr.Mummoorthy, A., Bhasker, B., & Jagadish Kumar, T. (2018). Using of Bellman Fords Algorithm in WSN to Identify the Shortest Path and Improve the Battery Power & Control the DDOS Attackers and Monitor the System Environment. *Bonfring International Journal of Networking Technologies and Applications*, 5(1), 9-11.
- [10] Andrea, and Matthew. (2017). FlybackSnubber to Recycle the Absorbed Energy in the Clamping Capacitor for Isolated Bidirectional Full Bridge DC-DC Converter. *Bonfring International Journal of Power Systems and Integrated Circuits*, 7(1), 19-25.
- [11] Neenu Preetam, I., & Gupta, H. (2014). Cardless Cash Access using Biometric ATM Security System. *International Scientific Journal on Science Engineering & Technology*, 17(10), 893-897.
- [12] Jha, H.R., & Singh, S.N. (2015). Study of Scattering Parameters and Gain of two Longitudinal Slots of Same Electrical Lengths Milled on two Waveguides for Series and Shunt Slot Array Planar Antenna. *Bonfring International Journal of Research in Communication Engineering*, 5(3), 12-21.
- [13] Mandal, S., Saha, G., & Pal, R.K. (2014). A Comparative Study on Disease Classification using Different Soft Computing Techniques. *The SIJ Transactions on Computer Science Engineering & its Applications*, 2(4), 1-29.
- [14] Gopalakrishnan, C., & Iyapparaja, M. (2019). Detection of Polycystic Ovary Syndrome from Ultrasound Images using SIFT Descriptors. *Bonfring International Journal of Software Engineering and Soft Computing*, 9(2), 26-30.
- [15] Sivasankari, M., & Dr.Velmani, P.& Dr.ArokiaJansi Rani, P. (2018). Multilingual Off-line Handwriting Recognition in Real-World Images Using Deep Neural Network (DNN) Classifier. *Journal of Computational Information Systems*, 14(4), 164 - 175.
- [16] Dr.Berin Jones, C. (2018). Text Segmentation and Recognition in Natural Scene Images Using MSER. *Journal of Computational Information Systems*, 14(5), 1 - 8.
- [17] Preethi, S., & Leelavathi, B. (2018). Adaptive Firefly Algorithm (AFA) based Feature Selection and Unsupervised Fuzzy Extreme Learning Machine (USUFELM) with Network-based Intrusion Detection and Prevention System. *Journal of Computational Information Systems*, 14(5), 34 - 44.
- [18] Khanaa, V., Thooyamani, K. P., & Udayakumar, R. (1798). Cognitive radio based network for ISM band real time embedded system. *Middle-East Journal of Scientific Research*, 16(12), 1798-1800.
- [19] Vijayaraghavan, K., Nalini, S.K., Prakash, N.U., & Madhankumar, D. (2012). Biomimetic synthesis of silver nanoparticles by aqueous extract of *Syzygium aromaticum*. *Materials Letters*, 75, 33-35
- [20] Caroline, M.L., Sankar, R., Indirani, R.M., & Vasudevan, S. (2009). Growth, optical, thermal and dielectric studies of an amino acid organic nonlinear optical material: l-Alanine. *Materials Chemistry and Physics*, 114(1), 490-494.
- [21] Kumaravel, A., & Pradeepa, R. (2013). Efficient molecule reduction for drug design by intelligent search methods. *International Journal of Pharma and Bio Sciences*, 4(2), B1023-B1029.
- [22] Kaviyarasu, K., Manikandan, E., Kennedy, J., Jayachandran, M., Ladchumananandasivam, R., De Gomes, U. U., & Maaza, M. (2016). Synthesis and characterization studies of NiO nanorods for enhancing solar cell efficiency using photon upconversion materials. *Ceramics International*, 42(7), 8385-8394.
- [23] Sengottuvel, P., Satishkumar, S., & Dinakaran, D. (2013). Optimization of multiple characteristics of EDM parameters based on desirability approach and fuzzy modeling. *Procedia Engineering*, 64, 1069-1078.

- [24] Anbuselvi S., Chellaram, C., Jonesh S., Jayanthi L., & Edward J.K.P. (2009). Bioactive potential of coral associated gastropod, *Trochus tentorium* of Gulf of Mannar, Southeastern India. *J. Med. Sci*, 9(5), 240-244.
- [25] Kaviyarasu, K., Ayeshamariam, A., Manikandan, E., Kennedy, J., Ladchumananandasivam, R., Gomes, U. U., & Maaza, M. (2016). Solution processing of CuSe quantum dots: Photocatalytic activity under RhB for UV and visible-light solar irradiation. *Materials Science and Engineering: B*, 210, 1-9.
- [26] Kumaravel, A., & Udayakumar, R. (2013). Web portal visits patterns predicted by intuitionistic fuzzy approach. *Indian Journal of Science and Technology*, 6(5S), 4549-4553.
- [27] Srinivasan, V., & Saravanan, T. (2013). Reformation and market design of power sector. *Middle-East Journal of Scientific Research*, 16(12), 1763-1767.
- [28] Kaviyarasu, K., Manikandan, E., Kennedy, J., & Maaza, M. (2015). A comparative study on the morphological features of highly ordered MgO: AgO nanocube arrays prepared via a hydrothermal method. *RSC Advances*, 5(100), 82421-82428.
- [29] Kumaravel, A., & Udhayakumarapandian, D. (2013). Construction of meta classifiers for apple scab infections. *International Journal of Pharma and Bio Sciences*, 4(4), B1207-B1213.
- [30] Sankari, S. L., Masthan, K. M. K., Babu, N. A., Bhattacharjee, T., & Elumalai, M. (2012). Apoptosis in cancer-an update. *Asian Pacific journal of cancer prevention*, 13(10), 4873-4878
- [31] Harish, B. N., & Menezes, G. A. (2011). Antimicrobial resistance in typhoidal salmonellae. *Indian journal of medical microbiology*, 29(3), 223-229.
- [32] Manikandan, A., Manikandan, E., Meenatchi, B., Vadivel, S., Jaganathan, S. K., Ladchumananandasivam, R., & Aanand, J. S. (2017). Rare earth element (REE) lanthanum doped zinc oxide (La: ZnO) nanomaterials: synthesis structural optical and antibacterial studies. *Journal of Alloys and Compounds*, 723, 1155-1161.
- [33] Caroline, M. L., & Vasudevan, S. (2008). Growth and characterization of an organic nonlinear optical material: L-alanine alaninium nitrate. *Materials Letters*, 62(15), 2245-2248.
- [34] Saravanan T., Srinivasan V., Udayakumar R. (2013). A approach for visualization of atherosclerosis in coronary artery, Middle - East Journal of Scientific Research, 18(12), 1713-1717.
- [35] Poongothai, S., Ilavarasan, R., & Karrunakaran, C.M. (2010). Simultaneous and accurate determination of vitamins B1, B6, B12 and alpha-lipoic acid in multivitamin capsule by reverse-phase high performance liquid chromatographic method. *International Journal of Pharmacy and Pharmaceutical Sciences*, 2(4), 133-139.
- [36] Udayakumar, R., Khanaa, V., & Saravanan, T. (2013). Synthesis and structural characterization of thin films of SnO<sub>2</sub> prepared by spray pyrolysis technique. *Indian Journal of Science and Technology*, 6(6), 4754-4757
- [37] Anbazhagan, R., Satheesh, B., & Gopalakrishnan, K. (2013). Mathematical modeling and simulation of modern cars in the role of stability analysis. *Indian Journal of Science and Technology*, 6(5S), 4633-4641.
- [38] Caroline, M.L., & Vasudevan, S. (2009). Growth and characterization of bis thiourea cadmium iodide: A semiorganic single crystal. *Materials Chemistry and Physics*, 113(2-3), 670-674.
- [39] Sharmila, S., Jeyanthi Rebecca, L., & Das, M. P. (2012). Production of Biodiesel from *Chaetomorpha antennina* and *Gracilaria corticata*. *Journal of Chemical and Pharmaceutical Research*, 4(11), 4870-4874.
- [40] Thooyamani, K.P., Khanaa, V., & Udayakumar, R. (2013). An integrated agent system for e-mail coordination using jade. *Indian Journal of Science and Technology*, 6(6), 4758-4761.
- [41] Caroline, M. L., Kandasamy, A., Mohan, R., & Vasudevan, S. (2009). Growth and characterization of dichlorobis l-proline Zn (II): A semiorganic nonlinear optical single crystal. *Journal of Crystal Growth*, 311(4), 1161-1165.
- [42] Caroline, M.L., & Vasudevan, S. (2009). Growth and characterization of L-phenylalanine nitric acid, a new organic nonlinear optical material. *Materials Letters*, 63(1), 41-44.
- [43] Kaviyarasu, K., Xolile Fuku, Gene T. Mola, E. Manikandan, J. Kennedy, and M. Maaza. Photoluminescence of well-aligned ZnO doped CeO<sub>2</sub> nanoplatelets by a solvothermal route. *Materials Letters*, 183(2016), 351-354.

- [44] Saravanan, T., & Saritha, G. (2013). Buck converter with a variable number of predictive current distributing method. *Indian Journal of Science and Technology*, 6(5S), 4583-4588.
- [45] Parthasarathy, R., Ilavarasan, R., & Karrunakaran, C. M. (2009). Antidiabetic activity of Thespesia Populnea bark and leaf extract against streptozotocin induced diabetic rats. *International Journal of PharmTech Research*, 1(4), 1069-1072.
- [46] Hanirex, D. K., & Kaliyamurthie, K. P. (2013). Multi-classification approach for detecting thyroid attacks. *International Journal of Pharma and Bio Sciences*, 4(3), B1246-B1251
- [47] Kandasamy, A., Mohan, R., Lydia Caroline, M., & Vasudevan, S. (2008). Nucleation kinetics, growth, solubility and dielectric studies of L-proline cadmium chloride monohydrate semi organic nonlinear optical single crystal. *Crystal Research and Technology: Journal of Experimental and Industrial Crystallography*, 43(2), 186-192.
- [48] Srinivasan, V., Saravanan, T., Udayakumar, R., & Saritha, G. (2013). Specific absorption rate in the cell phone user's head. *Middle-East Journal of Scientific Research*, 16(12), 1748-50.
- [49] Udayakumar R., Khanaa V., & Saravanan T. (2013). Chromatic dispersion compensation in optical fiber communication system and its simulation, *Indian Journal of Science and Technology*, 6(6), 4762-4766.
- [50] Vijayaragavan, S.P., Karthik, B., Kiran, T.V.U., & Sundar Raj, M. (1990). Robotic surveillance for patient care in hospitals. *Middle-East Journal of Scientific Research*, 16(12), 1820-1824.