A Novel RI-tree Based top-k subscription matching for location-aware Publish/Subscribe systems

Sushant Hulawale[1], Bharat Burghate[2]

[1] Student, Department of Computer Engineering, JSPM's, BSIOTR, Maharashtra, India
[2] Assistant Professor Department of Computer Engineering, JSPM's, BSIOTR, Maharashtra, India

ABSTRACT

The dissemination of messages to an enormous range of mobile users has raised lots of attention. This issue is inherent in rising applications, like location-based targeted advertising, information diffusive, and ride sharing. During this project, we have to examine a way to support location-based message dissemination in economically good and efficient manner. Our main plan is to develop a location-aware version of the Pub/Sub model that was designed for message dissemination. Whereas lots of studies have with success used this model to match the interest of subscriptions (e.g., the properties of potential customers) and events (e.g., data of casual users), the problems of incorporating the placement data of subscribers and publishers haven't been well addressed. We have propose to model subscriptions and events by Boolean expressions and site knowledge this permits complicated data to be given. However, since the amount of publishers and subscribers is huge, the time price for matching subscriptions and events is prohibitory. To deal with this drawback, we've got developed the RI-tree. This organization is an integration of the R-tree and also the dynamic interval-tree. Beside our novel pruning strategy on RI-tree, our resolution will effectively and with efficiency come back the top-k subscriptions with relevancy a happening. Compared to the prevailing works, the RI-tree projected during this project has 2 characteristic options. First, it permits users to specify their interests with Boolean expressions, that is additional communicative than keywords. Second, it focuses on the top-k semantics that is often utilized in several rising applications (e.g., location-based targeted advertising).

Keyword: *Online information services, web-based services, publishing.*

1. INTRODUCTION

Due to the advance of telecommunications and web technologies, tremendous amounts of location info will currently be obtained simply. As an example, a user's location is commonly half-tracked by base stations in an exceedingly cellular network; a vehicle's position are often obtained through GPS receivers or sensors on roads; a user reveals her location once she "checks in" (e.g., through Facebook and Twitter). the supply of location info stimulates the event of location- primarily based electronic communication services, that disseminates fascinating messages to users supported their positions and alternative info. Taking location-based targeted advertising as Associate in Nursing example, advertisements square measure sent to users hand-picked in terms ancient, gender, interest, and site one. Another example is that the Location-Based App Recommendation (LBAR), wherever code or "apps" square measure instructed to a user supported wherever she is. Associate in Nursing LBAR feature recently seems in Apple's iOS 8, that shows the image of Associate in Nursing app (e.g., "Starbucks") within the lock screen supported the user's context (e.g., she is near a Starbucks low shop). As recognized by vim Mobile in 2013, the LBTA outperforms non-location-targeted advertising by an element of 2, and therefore the usage of the LBTA exceeds the business average click-through rate (CTR) of 0.4%.

Those applications can be built on the top of Publish/Subscribe (Pub/Sub) systems which can provide large-scale matching and information dissemination. In a Pub/Sub system, there are two kinds of clients, subscriber and publisher. A subscriber, typically an information provider such as an advertising company, specifies the properties of users in which it is interested. These properties, or constraints, are collectively known as a subscription. For instance, an advertising company A (e.g., a restaurant), acting as a subscriber, posts the following subscription to the Pub/Sub system: (15 <age< 30, interest = {barbecue, sushi}, gender = male, visited time ≥ 3). The constraints specified in this subscription are used to match the "events" published by a publisher; once a matching is found, information from a subscriber is sent to the publisher. A publisher can be a casual mobile phone user. When a publisher, say, U, browses a homepage (say, Facebook), an event, containing information about this user (e.g., age=25, interest=barbecue, gender=male, visited time=5), is sent to the system.

Since one or more constraints could also be per a subscription, it should not be attainable for the values of a happening to match all the constraints. Hence, researchers have projected to permit a lot of flexibility within the matching method by permitting matching between subscribers and publishers to be inexact or partial. This variant of Pub/Sub systems, referred to as graded Pub/Sub systems, come back the k best subscriptions (or top-k subscriptions) to a publisher supported some marking functions. We notice that the present work solely focuses on traditional mathematical expressions together with strings and numbers. However, the freshly rising applications bring the new technical challenges. Continued the instance of a location-based targeted advertising application, once a publisher opens an app, his geographic coordinates are sent to the system. A subscriber may also specify this type of location, as an example, by voice communication that his look is at a particular location. On the opposite hand, to push advertisements to a mobile user, owing to the factors like restricted network information measure and also the screen size of user's movable, solely the advertisements whose distribution scopes are just about the user's current location could become the candidate advertisements. Since subscriptions contain each complicated mathematical expressions and site info, it's rather pricey to retrieve top- k relevant subscriptions from scores of subscriptions for a happening. Different from the graded Pub/Sub systems, the placement info has been integrated into the alleged keyword mathematical matching Pub/Sub systems wherever the numeric attribute matching isn't supported. Their ways of incorporating the placement info are either not applicable or inefficient for the graded Pub/Sub systems. Within the empirical studies, we have extend the present work associated with the location-aware Pub/Sub systems because the competitors to support the numeric attribute matching. The experimental results demonstrate that our approach considerably outperforms the competitors. During this paper, we have to explore the problems of incorporating location info into a graded Pub/Sub system. To support top-k subscription matching for location-aware Pub/Sub systems, we have a tendency to propose a unique R-tree based mostly index, the RI-tree, by group action the dynamic interval tree into the R-tree nodes. Once a happening with location info arrives, our formula will quickly report the top-k subscriptions most relevant to the event. To summarize, our main contributions are:

1. We formalize a new variant of top-k subscription matching, permitting location data to be a part of a subscription or an event.

2. We propose an index structure, called the RI-tree.

3. We design an efficient matching algorithm.

2. LITERATURE SURVEY

2.1 An efficient spatial publish/subscribe system for intelligent location-based services

The advance in wireless web and mobile computing brought the booming of intelligent Location-Based Services (LBS), which may actively push location-dependent info to mobile users in line with their predefined interest. The successful development of push-based LBS applications depends on the existence of a publish/subscribe middleware which will handle spatial relationship. This paper presents economical spatial publish/subscribe system which will function the middleware for intelligent LBS applications. The fundamental models, together with spatial event model, spatial subscription model and notification model, are introduced and therefore the over-all design is presented. Two forms of spatial predicate which will meet most typical requirement of intelligent location aware applications also are mentioned. What is more, we have propose novel spatial event process approach that dispatches the spatial subscriptions to self-positioning mobile devices. By investing client-side computing resource and decreasing the communication times, the server-side load is eased and therefore the communication price is reduced [1].

2.2 Semi-Probabilistic Content-Based Publish-Subscribe

Mainstream approaches to content-based distributed publish-subscribe usually route events deterministically supported data collected from subscribers, and do so by relying on a tree-shaped overlay network. Whereas this answer achieves scalability in fixed, large-scale settings, it's less appealing in situations characterized by high dynamicity, e.g., mobile ad hoc networks or peer-to-peer systems. At the other extreme, researchers within the connected fields of multicast and cluster communication have with success exploited probabilistic techniques that give raised fault tolerance, resilience to changes, and however are scalable. In this paper, we tend to propose a completely unique approach wherever event routing deterministic decisions driven by a restricted view on the subscription info and, if this is often not sufficient, resorts to probabilistic choices performed by choosing links random [3].

2.3 On introducing location awareness in publish-subscribe middleware

Having the likelihood of routing messages solely toward specific areas or subscribing to messages originating in specific locations looks natural once a publish-subscribe model of communication is adopted. Sadly, only a few work have investigated such a types services and none of the foremost widely adopted publish-subscribe middleware implements them. In this paper we have to initially classify possible location-based publish-subscribe services, then we have to describe an algorithmic program which efficiently implement them in a distributed publish-subscribe middleware [4] [5].

2.4 Pervaho: A specialized middleware for mobile context-aware applications

The idea of context-awareness offers a good potential for future of mobile applications. In order to be developed in an optimal manner, mobile context-aware applications require acceptable middleware services. This paper introduces Pervaho, an integrated middleware aimed specifically at supporting the development and testing of mobile context-aware applications. As an example the utilization of Pervaho, we go through the development of a concrete mobile application and show however it will be engineered on top of Pervaho's location-based publish/subscribe service. We also illustrate however a specialized quality testing tool considerably simplifies the method of testing proximity-based semantics. We then present the implementation of Pervaho, that is depending on a set of communication protocols geared at mesh networks [6].

2.5 Supporting mobility in content-based publish/subscribe middleware

Publish/subscribe (pub/sub) is taken into account a valuable middleware architecture that proliferates loose coupling and leverages re-configurability and evolution. Up to now, existing pub/sub middleware was optimized for static systems where users as well as the underlying system structure were rather fixed. We have to study the question whether or not existing pub/sub middleware are often extended to support mobile and location-dependent applications. We 1st analyze the wants of such applications and distinguish two orthogonal forms of mobility: the system-centric physical mobility and an application-centric logical mobility (where users are aware that they're changing location). We need introduce location-dependent subscriptions as an acceptable means that to exploit the power of the event-based paradigm in mobile applications. In short spoken, location-dependency refines a subscription to simply accept only events associated with a mobile user's current location [7].

2.6 Context-aware broadcasting approaches in mobile ad hoc networks

The aim of this paper is to check totally different context-aware broadcasting approaches in mobile ad hoc networks (MANETs) and to evaluate their respective performances. Message broadcasting is one among the core challenges named by distributed systems and has so for the most part been studied within the context of ancient network structures, like the Internet. With the emergence of MANETs, new broadcasting algorithms particularly geared at these networks are introduced. The goal of those broadcasting algorithms is to confirm that a most range of nodes deliver the broadcasted message (reliability), whereas guaranteeing that the minimum range of nodes retransmit the broadcasted message (efficiency), so as to avoid wasting their resources, like bandwidth or battery. In recent years, as more and more mobile devices became context-aware, many broadcasting algorithms are introduced that make the most of contextual information so as to boost their performance. We have distinguish four approaches with relevancy context:
(i) Context-oblivious approaches, (ii) Network traffic-aware approaches, (iii) Power-aware approaches, and (iv) Location-aware approaches [7].

2.7 Retrieving top-k prestige-based relevant spatial web objects

The location-aware keyword query returns graded objects that are close to a query location which have textual descriptions that match query keywords. This query happens inherently in many sorts of mobile and traditional internet services and applications, e.g., phone book and Maps services. Previous work considers the potential results of such a query as being independent once ranking them. However, a relevant result object with close objects that also are relevant to the query is probably going to be desirable over a relevant object while not relevant close objects. The paper proposes the concept of prestige-based relevancy to capture each the textual relevancy of an object to a query and therefore the effects of close objects. On this, a new variety of queries, the Location-aware top-k Prestige-based Text retrieval (LkPT) query, is planned that retrieves the top-k spatial internet objects graded in both prestige-based relevance and location proximity. We propose two algorithms that compute LkPT queries [8].

2.8 Collective Spatial Keyword Querying

With the proliferation of geo-positioning and geo-tagging, spatial internet objects that possess each a geographical location and a textual description are gaining in prevalence, and spatial keyword queries that exploit each location and textual description are gaining in prominence. However, the queries studied to this point usually focused on finding individual objects that each satisfy a query instead of finding groups of objects wherever the objects in a groups collectively satisfy a query. We outline the problem of retrieving a group of spatial internet objects specified the group's keywords cover the query's keywords and specified objects are nearest to the query location and have all-time low inter-object distances. Specifically, we study two variants of this downside, each of that is NP-complete. We devise actual solutions furthermore as approximate solutions with obvious approximation bounds to the issues [8] [9].

2.8 Efficient query processing in geographic web search engines

Geographic internet search engines enable users to constrain and order search leads to an intuitive manner by focusing a query on a selected geographical region. Geographic search technology, also known as local search, has recently received important interest from major companies of search engines. Tutorial analysis during this space has focused totally on techniques for extracting geographic information from the online. In this paper, we study the problem of economical query processing in scalable geographic search engines. Query processing may be a major bottleneck in normal internet search engines, and also the main reason for the thousands of machines utilized by the main engines. Geographic search engine query processing is completely different in this it needs a mix of text and spatial data processing techniques. We propose many algorithms for economical query processing in geographic search engines, integrate them into an existing internet search query processor, and evaluate them on giant sets of real information and query traces [9].

3. PROPOSED SYSTEM

Location-Aware Pub/Sub Systems

Recently, there are several researches on location-aware Pub/Sub systems from a database perspective. G., Wang proposes the Rt-tree which may expeditiously filter geo-textual information. M., Li extends Rt-tree to support ranking semantics, i.e., return all subscriptions whose similarities with the query event aren't smaller than a given threshold θ. However the pruning algorithm cannot be used for top-k search. Further, studies the location-aware Pub/Sub problem for parameterized spatio-textual subscriptions and presents a filter-verification framework by integrating prefix filtering and spatial pruning techniques [10]. However, only keywords are considered, and that they cannot support to retrieve top-k subscription matching. Note that Pub/Sub systems that only think about keywords cannot support numeric attribute matching like the instance of the subscription. Compared to existing works, the RI-tree planned in this paper has two characteristic options. First, it permits users to specify their interests with boolean expressions, that is a lot more communicative than keywords. Second, it focuses on the top-k semantics that is often utilized in several rising applications (e.g., location-based targeted advertising). Chen et al. considers the temporal spatial-keyword top-k subscription query. They present an economical solution which may incessantly maintain up-to- date top- k most relevant results (events) over a stream of geo-textual objects for every subscription. Guo, L., Zhang proposes new location-aware Pub/Sub system, i.e., Elaps, that focuses on incessantly watching moving users subscribing to dynamic event streams. However, their issues are completely different from ours. Our work is completely different from spatial keyword search. The main reason is that they focus on keywords whereas we to adopt boolean expressions to specific subscriber's needs in subscriptions, that is a lot more communicative that keywords.Given a collection of subscriptions S, an event e, and a parameter k, the Top-k Subscription Matching problem (SM-k problem for short) finds the top-k best matching set.In this project, we present a framework that integrates the R-tree and also the interval-tree into a new index, named RI-tree which includes an algorithm for process SM-k problem using the RI-tree.

RI-tree Index Structure

The R-tree could be a widely used index for spatial queries and therefore the interval-tree is that the "standard" better-known simple processing straightforward stabbing queries. They're designed on an individual basis for various forms of queries. The RI-tree is actually R-tree, every node of that is enriched with regard to a group of dynamic interval trees for objects contained in its sub-tree. In the RI-tree, if node N could be a leaf node, it contains variety of entries of forms (sid, Ω, loc, α), wherever sid is that the symbol of subscription, Ω, loc and α are the Boolean expression, the location and tuning parameter of the subscription s sid, respectively. Here, it's vital to notice that the weight of a subscription ssid on every attribute is multiplied by (1 − ssid.α) once ssid is inserted into the RI-tree. A leaf node also contains some metadata. The metadata includes rectangle, that is the Minimum Bounding rectangle of all constituent entries, α min and α max that are the minimum and maximum values of α among all constituent entries, and therefore the aggregated information Γ for every attribute of the form (attr i, range, ω max). Additionally, a leaf node also contains a pointer to a dynamic interval tree forest F, i.e., a set of dynamic interval trees organized by a hashmap. Let S attr be the set of all attributes stored on the leaf node. The hashmap manages all attributes in S attr. For every attribute attr i, the hashmap maps it to a dynamic interval tree T attr i. The tree T attr i stores all intervals of the leaf node N's entries on attribute attr i. The form of intervals stored on the tree is (range, sid, ω), wherever range denotes the range of the interval, sid is that the id of the corresponding subscription and ω is that the weight of that subscription on attr i. The dynamic interval tree T attr i dynamically maintains a set of intervals I, wherever every interval I ∈I has a weight I. ω such the interval with the max weight containing an event point can be found efficiently. This structure will solve the interval stabbing-max problem. . For instance, for the query point q, it stabs four intervals (a, b, c, d). Since interval d has the greatest weight 0.5, it will be returned. There exists several solutions which can solve this problem. In this paper, we use the modified interval tree structure which can support queries in O(log 2 n) time, updates in O (log n) time and only requires O(n)space. On the opposite hand, if node N could be a non-leaf node, it contains variety of entries cp, that points to the corresponding child node. Being same type form of leaf nodes, node N maintains the data and a dynamic interval tree forest organized by a two-level index structure that contains the interval data for every associated attributes. For every child node U of node N, the interval of U. Γ attr i will be store on the dynamic interval tree T attr i of the node N. Thus, the amount of intervals in T attr i will not exceed the amount of entries in node N.

4. LOCATION-AWARE PUB/SUB SYSTEM ARCHITECTURE

Location aware publish/subscribe system consists of three main components: Publisher, Server and Subscriber. The communication between these components requires registration to the server. Publisher mainly generate the event and register its details to the server like the area of interest, location information.

Once the event is created on the server, then server starts publishing the event to the subscription matching subscribers according to the location for which the event is applicable.

But for this subscriber also has to be registered on the server with his details and area of interest.

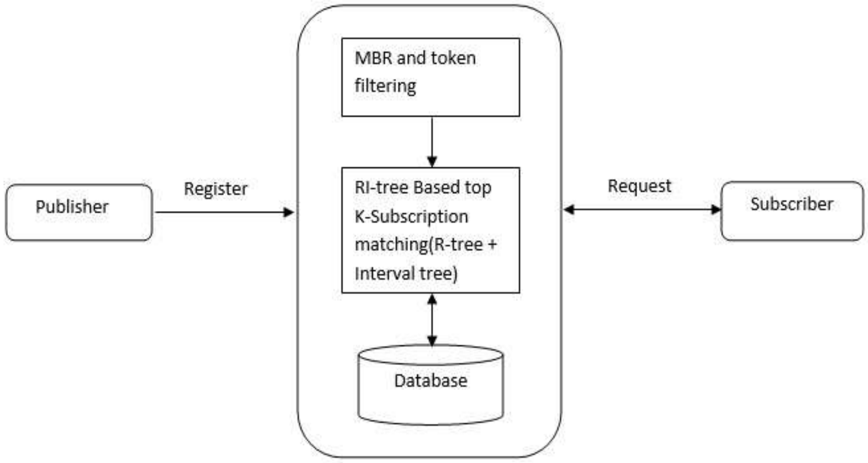


Fig -1 System Architecture

4.1 Publishers & Subscribers Register with Pub/Sub Server

Publish–subscribe is a messaging pattern where senders of messages, called publishers, do not program the messages to be sent directly to specific receivers, called subscribers, but instead characterize published messages into classes without knowledge of which subscribers, if any, there may be.

Similarly, subscribers express interest in one or more classes and only receive messages that are of interest, without knowledge of which publishers, if any, there are.

Pub/sub Server is a sibling of the message queue paradigm, and is typically one part of a larger message-oriented middleware system. Most messaging systems support both the pub/sub and message queue models in their API, e.g. Java Message Service (JMS).

In this module both publisher and subscribers are register with Pub/Sub Server.

4.2 Adaptive Subscriber's Query Search

In this module subscriber's can subscribe for interest domain. In this project we use Push Model. It means whenever related data published by publisher will automatically reach subscriber.

Existing methods are pull method. It means subscriber request every time to get related data.

We are considering location of publisher and subscriber.

In this module subscribers register subscriptions to capture their interests. A subscription s includes a textual description s. T and spatial information s. R, denoted by s = (T, R).

The spatial information is used to capture a subscriber's most interested region. We use the well-known minimum bounding rectangle (MBR) to denote a region s. R.

The textual description is used to capture a subscriber's content-based interests, denoted by a set of tokens s. T = {t1, t2,...,ts,T}.

4.3 Publish contents

In this module, publisher can publish any content with any domain.

A message m posted by a publisher also contains a textual description m.T and spatial information m.R, denoted by m = (T, R), which respectively have the same meaning as those of subscriptions.

Note that the spatial information m.R of a message can be a point, e.g., mobile user's location. If the spatial information of a message is a point, we call it point message; otherwise we call it range message.

4.4 RI-tree Based top-k subscription matching

In this module takes a subscription si is an element of S (S={s1,s2,...,sS}) and a message m, a location-aware publish/subscribe server delivers m to si (si is called an answer of m) to subscribers.

This data structure is an integration of the R-tree and the dynamic interval-tree. Together with our novel pruning strategy on RI-tree, our solution can effectively and efficiently return the top-k subscriptions with respect to an event.

5. MATHEMATICAL MODEL

Let S, be a system such that,
S = {s,Subi,Pubi,Si,E}
Where,
S- Proposed System
s- Initial state at Time T = 0.
Subi = the Registered Subscriber with attributes {Sid,pwd,Li} where
Sid = subscriber_id
Li = Location_id
Pubi = the Registered Subscriber with attributes {Pid,pwd,Li} where
Pid = subscriber_id
Li = Location_id
Si = the Subscription that registers Subi i.e subscriber's interest {si,li}
Si= subscription_id
Li = Location_id
E = events that are published by Pubi i.e publisher {Ei,li}
Ei = event_id
Li = location_id

E – Event
N – Node
S – Subscription
UB – Upper Bound
Sid – Identifier an subscription

X- Input of System.
-Si(subcription),E(event) and R tree node
Y- Output of System.
- Top K Subscription matching

T- Set of steps to be performed for Subscription matching
Process= {si, E}
1) E is parsed to get location li information
2) Once location range is confirmed Interval tree is generated for that interval/range
3) Top K subscriptions are matched within location range
4) Return top K subscription

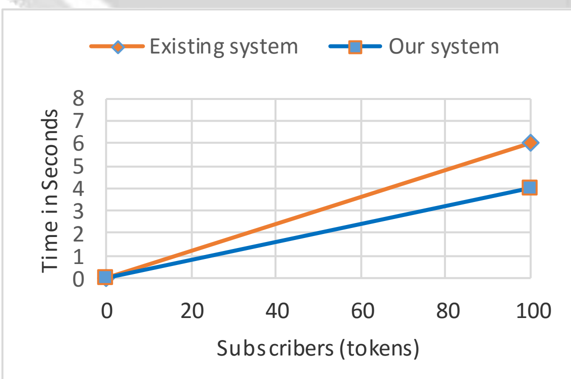
6. PERFORMANCE ANALYSIS

We have generated 3 groups of messages as follows:

(1) Short Point Messages: Each message contained 6-20 tokens and had a point location.

(2) Long Point Messages: Each message contained 100-1000 tokens and had a point location.

(3) Short Range Messages: Each message contained 6-20 tokens and had an MBR region.



For evaluating the performance of system we consider here the Length of Messages generated for a event by a publisher vs time required for sending it to the subscriber in locality.

For Existing system for sending the Messages consisting of around 100 tokens require 6 seconds. For this system for sending the Messages consisting of around 100 tokens is expected to require 4 seconds.

7. CONCLUSION

In this paper, we propose and formalize a variant of top-k subscription matching, i.e., top- k subscription matching for location-aware Pub/Sub systems which supports boolean expressions in subscriptions. We propose a novel index structure RI-tree, which combines the R-tree and the dynamic interval-tree. In addition, we develop an efficient filtering strategy to reduce the search space. Finally, we evaluate the RI-tree based solution by experiments on a large-scale dataset. Processing/communication overheads are high in our proposed work. Because users have to continuously issue queries.

8. REFERENCES

[1] X. Chen, Y. Chen, and F. Rao, "An efficient spatial publish/ subscribe system for intelligent location-based services," in Proc. 2nd Int. Workshop Distrib. Event-Based Syst., 2003, pp. 1–6.

[2] L. Fiege, F. C. G €artner, O. Kasten, and A. Zeidler, "Supporting mobility in content-based publish/subscribe middleware," in Proc. USENIX Int. Conf. Middleware, 2003, pp. 103–122.

[3] P. Costa and G. P. Picco, "Semi-probabilistic content-based publish-subscribe," in Proc. 25th IEEE Int. Conf. Distrib. Comput. Syst., 2005, pp. 575–585.

[4] G. Cugola and J. E. M. de Cote, "On introducing location aware-ness in publish-subscribe middleware," in Proc. IEEE Int. Conf. Distrib. Comput. Syst. Workshops, 2005, pp. 377–382.

[5] P. T. Eugster, B. Garbinato, and A. Holzer, "Location-based publish/subscribe," in Proc. 4th IEEE Int. Symp. Netw. Comput. Appl., 2005, pp. 279–282.

[6] P. T. Eugster, B. Garbinato, and A. Holzer, "Pervaho: A specialized middleware for mobile context-aware applications," Electron. Commerce Res., vol. 9, no. 4, pp. 245–268, 2009.

[7] B. Garbinato, A. Holzer, and F. Vessaz, "Context-aware broadcast-ing approaches in mobile ad hoc networks," Comput. Netw., vol. 54, no. 7, pp. 1210–1228, 2010.

[8] X. Cao, G. Cong, and C. S. Jensen, "Retrieving top-k prestige-based relevant spatial web objects," Proc. VLDB Endowment, vol. 3, no. 1, pp. 373–384, 2010.

[9] Minge Yu, Guoliang Li, Ting Wang, Jianhua Feng, Zhiguo Gong "Efficient Filtering Algorithms for Location-Aware Publish/Subscribe" in IEEE vol.27 no.4, 2015