Mobile Information Systems 7 (2011) 107–122
DOI 10.3233/MIS-2011-0113
IOS Press

Semantic file annotation and retrieval on mobile devices

Sadaqat Jan*, Maozhen Li, Ghaidaa Al-Sultany, Hamed Al-Raweshidy and Ibrar Ali Shah
School of Engineering and Design, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK

Abstract. The rapid development of mobile technologies has facilitated users to generate and store files on mobile devices such as mobile phones and PDAs. However, it has become a challenging issue for users to efficiently and effectively search for files of interest in a mobile environment involving a large number of mobile nodes. This paper presents SemFARM framework which facilitates users to publish, annotate and retrieve files which are geographically distributed in a mobile network enabled by Bluetooth. The SemFARM framework is built on semantic web technologies in support of file retrieval on low-end mobile devices. A generic ontology is developed which defines a number of keywords, their possible domains and properties. Based on semantic reasoning, similarity degrees are computed to match user queries with published file descriptions. The SemFARM prototype is implemented using the Java mobile platform (J2ME). The performance of SemFARM is evaluated from a number of aspects in comparison with traditional mobile file systems and enhanced alternatives. Experimental results are encouraging showing the effectiveness of SemFARM in file retrieval. We can conclude that the use of semantic web technologies have facilitated file retrieval in mobile computing environments maximizing user satisfaction in searching for files of interest.

Keywords: Mobile computing, semantic, pervasive, ontology, file retrieval, SemFARM, ubiquitous

1. Introduction

Mobile phone usage has grown exponentially in recent years and became an essential part of our daily life. Mobile phones are not only used for communication purposes, their usage can also be for our daily health, business and entertainment purposes. Its usage in the field of education is prevalent [40] and different aspects of its importance are evident in world community. Once users start using these devices and their applications, it generates a large number of files. Although these hand held devices are limited in computational resources they are still commonly available with larger memory sizes to store audio, video, image, text and other types of files. These files are stored with application specific default naming settings; for example, a mobile phone camera generates an image file and usually names it as image001 and video recorder names its file as video001 which are non descriptive. These files are usually stored in hierarchical directory structures and then searched through browsing and navigated through these directories. This file retrieval approach becomes less affective and more time consuming as the number of files grow on devices with limited user interface capabilities. As a result, users will no longer be able to easily know where to find the required files and must browse all the directories and open files to find their contents. The issue can be dealt with by annotating files with associative tags and user-entered keywords which can later be used for searching. Keyword-based file searches have proven to be more

*Corresponding author: Sadaqat Jan, 258-Y, Michael Sterling, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. Tel.: +44 0 7590110862; E-mail: Sadaqat.jan@brunel.ac.uk, Sadaqat_jan@hotmail.com.

efficient only if a user remembers any associated keyword which is highly unlikely when the number of files on a storage device grows.

However, in similar situations, semantic technologies can play a vital role and offers a feasible approach to knowledge management and information processing. In this case, it makes it possible to retrieve a file even if a user does not exactly remember any associated keyword. Inspired by semantic web, we extended the approach to manage and retrieve files and documents on local devices and specifically on low-end devices where resources are limited especially in terms of input (keypads) and output (smaller screen sizes).

This paper presents a framework named SemFARM, which provides semantic based search by implementing semantic web technologies for use on low-end devices. If a user does not type even an exact keyword, SemFARM still might be able to retrieve the required file. Ontology can be used to deal with the management of keywords which are associated with files stored on a device. A conceptualization of a domain related to relations and concepts can be defined by an ontology to efficiently manage and process the knowledge, as advocated and explained in [14]. For this purpose, a general ontology is developed which concerns general common keywords, their domains and properties using Web Ontology Language (OWL) [17]. The search module of SemFARM dynamically creates a Resource Description Framework (RDF) schema extracted from an Extensible Mark-up Language (XML) file which contains meta-data of the files of a device. In this way the structure of the XML file is used to define the relationship amongst different tags. The XML file is created in an annotation process and explained later in Section 4.1. The RDF schema and ontology definitions are bind to get inferred information from the reasoner used in SemFARM. The search module sends results back to the query requesting device after completing the search process on a network server as explained in Section 5.2. SemFARM is also capable of searching a particular file of interest in networked environment when devices are connected through Bluetooth. Recent advancements in mobile technology enabled its users to share several types of contents through various interfaces in a range of environments. Research and development in the field of mobile platforms and software are equally growing for diverse applications including operating systems, user interfaces and audio/video applications in standalone or networked environment. Some efforts can also be seen to enhance video streaming in Vehicular Ad Hoc Networks where various obstacles arise due to multi-hop routing and mobility patterns [35]. In this paper, our main focus is to deal with the file on low-end devices and their efficient retrieval.

The rest of the paper is structured as follows. Section 2 presents the relevant research work. Section 3 gives an overview of kXML parser. An introduction to FARM is presented in Section 4. SemFARM implementation, its working model and its use in a case study are explained in Section 5. Performance evaluation is presented in Section 6, and finally, Section 7 concludes the paper, followed by the references.

2. Related work

Scientists and researchers have adopted and proposed several approaches towards efficient file retrieval. Most of these efforts aim at the retrieval of documents/WebPages on the internet because of the larger volume of information [30,39]; however, as the information volume increases on local resources like the Personal Area Network (PAN), desktop computers and mobile phones, efforts can now be seen towards efficient information retrieval on all such environments. For example, a semantic file system is proposed for file retrieval using virtual directories and extendable UNIX based file system integrating search functions [37]. Similarly, for supporting semantics in file systems, TagFS is proposed which allows file tagging and the tag-based browsing of information objects on top of an underlying file system [42]. Information retrieval becomes more challenging task as the restrictions increases in any mobile computing paradigm. A.B. Waluyo et al. presented a survey in which they differentiate the query optimization and processing mechanisms in mobile databases [4] and presented a state-of-the-art in data management for location-dependent query and processing techniques [5].

Various studies show the importance of tagging [32], concluding that annotation makes the retrieval more efficient, not only for images and videos but for any type of information including files retrieval. Flickr [15] is a special purpose web service for sharing user uploaded photos and ZoneTag [48] is a tool to annotate camera photos. The ZoneTag mobile application, which is also supported and analyzed by Naaman et al. [34], suggests context based tags and some additional tags when a photo is take on mobile phone camera. The importance of tags and annotation can be determined when retrieving a required photo on Flickr where a photo with more tags can easily be retrieved compare to a photo with fewer tags. Furthermore, another approach was proposed by Karypidis et al. [1] to annotate photos taken by mobile phone cameras by adding contextual information to them and devices were allowed in PAN to maintain a shared perception regarding the context to annotate files. The context information was stored on a common repository with the file annotation process being automatic. In order to utilize tagging in image retrieval , A. Wilhelm et al. [8] proposed a system to annotate images at the capture time by adding Phone ID, username, date and time. Similarly, a framework was described by Monaghan et al. [31] to use web services, sensors and ontologies to create meaningful annotations.

The semantic approach is also extended to mobile devices for picture retrieval, where pictures are annotated with contextual information and used to index each of them [47]. Similarly, the contextual ontology was introduced and successfully implemented in several research efforts for example, context ontology for mobile devices was developed from embedded mobile sensors [2] for using the resources efficiently, the FLAME2008 platform was successfully developed to support mobile users with personalized context-aware services [33], the context ontology was used in a prototype to supervise the health condition of elderly people in runtime [41]. Iwamoto et al. proposed a design called *uPhoto*, in which context based annotation was implemented by extracting information automatically from embedded sensors and use them as image annotation [44]. However, to the best of our knowledge no such real efforts have been made to annotate or develop ontology for all types of stored files on a mobile phone or other hand held device. Semantic technologies are used in several research studies supporting pervasive and ubiquities mobile computing. For example, Izumi et al. [18] examined the design of social context-awareness ontology for their implementation of a prototype to supervise elder people in a ubiquitous computing environment and Guo et al. [16] used ontology for dealing with objects in order to search physical artefacts and detect hidden objects in a smart indoor environment.

3. kXML

kXML is a lighter compact version of XML parser, specifically designed for low-end devices, and is exclusively used on the J2ME platform. XML is a meta markup language which was endorsed by W3C [46] and became a universally supported specification for exchanging documents and data across applications and platforms [13]. It has a standard syntax for meta-data and a standard structure for document and data. The human readable plaintext form of XML makes it application independent and readable to everyone. In addition, it provides a simple and standard syntax for encoding [25]. XML documents need to be accessed and manipulated by a processor called XML parsers which tends to be bulky and requires heavy runtime memory. User Interface Markup Language (UIML) is a dialect of XML language that allows expressing user interface and abstracts it generating user interfaces for

Fig. 1. FARM main menu.

platforms, like PDA etc. [3]. Further down the course A. Santangelo et al. proposed the eXtensible Presentation Language (XPL) which is based on design pattern paradigm that differentiate keep separate the presentation layer from the programming logic [9].

kXML is widely used pull parser adopting the Mobile Information Device Profile (MIDP) requirements [23]. There are three types of parsers which include model, push and pull. Model parsers create a representation of the whole document after reading it, and hence, require more memory than the other types of parsers [36]. Push parsers always process data definitions before the document and a complete tree structure is created in the memory. The generation of this tree is memory expansive and thus makes this category unsuitable for low-end devices. On other the hand, pull parsers read the data first before parsing it from definition. These parsers structure the document tree using recursive functions.

4. FARM implementation

The FARM framework, which is implemented in the J2ME platform, consists of several MIDlets that support file annotation and search functionalities [20]. The framework automatically annotates each file with three basic attributes in addition to two optional keywords. The search module provides various search functionalities on a device itself or on other connected devices. The Bluetooth module is provided and used when a file is intended to be searched or transferred between connected devices. The FARM search-module match keywords associated with file, therefore, an exact match is required to retrieve a file. The main menu of the application MIDlet developed for FARM can be seen in Fig. 1 which shows various options including file sharing, transfer, annotation and search options.

4.1. Annotation process

The Annotation process automatically traverses the directories for stored files, and for each file three basic attributes are extracted from the underlying operating system of the device. Figure 2 explains the logical work flow of the annotation process in FARM and the same process is used in SemFARM also.

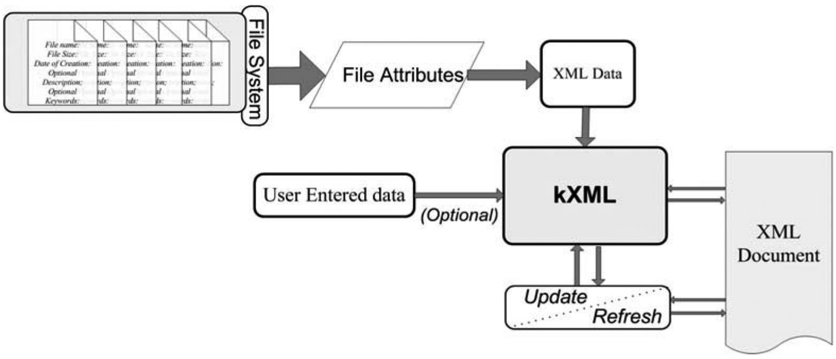


Fig. 2. Annotation process.

Three attributes, which include file-name, file-size and date-of-creation are then parsed through kXML and stored in an XML file locally. In addition to automatically created three attributes, the framework also allows to associate two optional keywords with each file. The *Update/Refresh* module is used to recreate the XML document when the whole meta-data is intended to be refreshed. This module is also used to update a single or multiple attributes of an individual or multiple files by parsing the XML document through kXML. As an example, the attributes of a file named "*14.txt*" are shown in Fig. 3a, and the same meta-data can be seen as a view option in Fig. 3b. The use of kXML parser allows the MIDlet to easily process the XML file for an update or searching for a specific tag on a resource limited device. Once the XML file is created, MIDlet also provides functionalities to edit, refresh or view complete XML file as a continuous list on screen. Users can edit the meta-data of any single file, or any single field, as shown in Fig. 3c and 3d respectively.

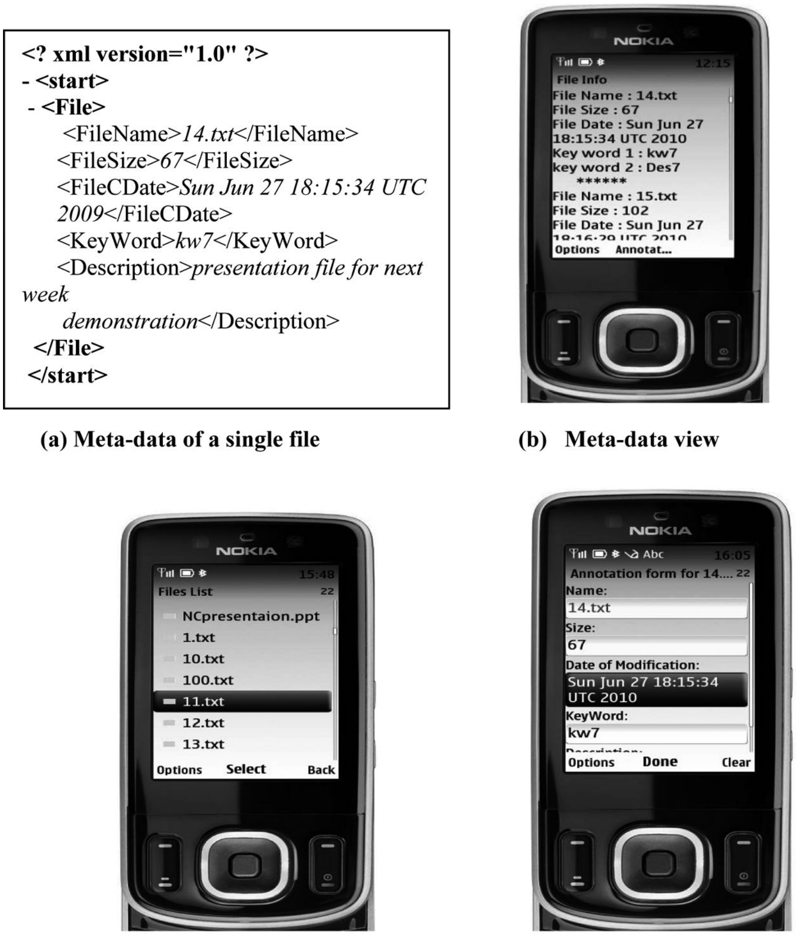
4.2. File search in FARM

A file search in FARM is performed with different functionalities which include, viewing a whole list of meta-data of all files on a device, using advance options to search by a specific field, or sending queries to all devices connected through Bluetooth. All search options make use of kXML parser to parse the XML files and search for a specific field match. If the search is requested from a nearby connected device, the result is sent back to the requesting device.

5. SemFARM implementation

SemFARM is built in J2ME platform which facilitates file annotation, sharing and semantic based file search. SemFARM uses the same annotation process explained in Section 4.1 where user can associate and edit metadata of files.

Several implementations and context-aware systems have been developed through Semantic Web technologies [45], such as ontology, RDF and OWL [27,38]. RDF is a standard model for data interchange which is widely used to share and communicate ontology and it also offers common properties and syntax for describing information. OWL was designed to be used by applications which need to process information contents and represent its machine interpretable contents on the web. Comparatively, OWL also adds more vocabulary with formal semantics and allows more expressive power. OWL itself is an evolution of DAML+OIL [28] which is divided into three sub-languages: OWL-Lite, which



(a) Meta-data of a single file **(b) Meta-data view**

(c) File selection for annotation **(d) Annotating fields**

Fig. 3. Annotation functionalities.

provides a hierarchy of classification and constraints; OWL-DL which has maximum expressiveness with computational completeness; and OWL Full which gives maximum expressiveness with-out a computational guarantee. The jena2 [19,21] toolkit provides the ability to parse and performs reasoning based on real standards, and it has been implemented in SemFARM. It is a leading toolkit for Java programmers in semantic web [10] and gives access to a range of inference capabilities. The reasoning subsystem of Jena2 allows various inference engines to be plugged-in which are used to derive additional information from base RDF combined with ontology definitions. The types of inference can be divided into two main types: standard and rule based. Standard inference includes RDFS and OWL reasoners while in rule based inference, jena allows programmers to define their own rules using jena APIs. The Jena2 inference structure [22], shown in Fig. 4, explains that the reasoner is accessed through a model factory to associate data, developing a new model which is called an inference model.

The collection of RDF statements, sometimes referred to as graphs, are associated with ontology definitions which gives additional statements that cannot directly be derived from RDF alone. We used the OWL reasoner which binds our ontology definitions with XML metadata dynamically converted to the RDF schema, as explained in the following section.

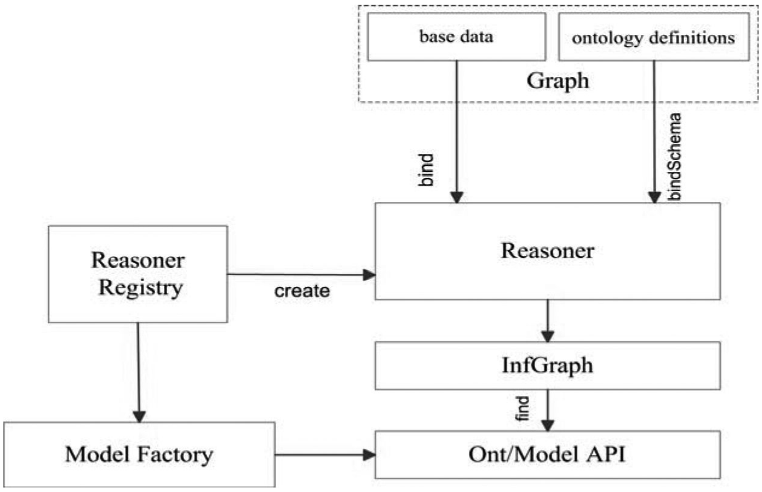


Fig. 4. Jena inference structure.

5.1. Use case study of SemFARM

Files are usually stored on mobile devices with the application default settings which are not descriptive enough to be used for file retrieval. The case worsens if users have larger storage capacities on their devices, which is very likely. A similar scenario is presented where a mobile phone user, assuming his name is *Michael,* took a few snapshots of family members using his mobile phone's built-in camera on a birthday party for his niece. On the same occasion, his wife and son also took snapshots using their own mobile phones.

Four months later, Michael wanted to view one of the group pictures taken at the party but he forgot the file name as the pictures were stored with application default name settings. Michael had to browse and view all the stored pictures on his mobile phone with 16GB of memory making his job more tedious, particularly on a limited keypad and screen. SemFARM can facilitate users in similar situations by providing various search options. It is expected that a user can even forget keywords associated with a file; but still the semantic support enables the framework to successfully accomplish the retrieval. User can utilize the search options of SemFARM by simply entering the date, the keyword associated with that particular file or any similar keyword for example *birthday*, *birthdayparty*, *party*, *niece* etc. The SemFARM will first search those files which have exact matches and then the ontology will be used to find similar keywords or meaning of the keyword which the user has entered to find the required file. If Micheal's phone is connected with mobile phones of his wife and son using Bluetooth, SemFARM will automatically perform the same search operation simultaneously on all mobile phones provided his son and wife have authenticated the connection and operation.

5.2. SemFARM search module

To perform a file search, the XML data created in the annotation process is sent to the search agent where the file is parsed to create an RDF document dynamically. In the creation of RDF model process, the XML structure is strategically dealt and exploited in such a way that tags associated with a file, are used as relationships in the RDF schema for a particular file. The RDF schema is then passed to the search Jena Inference engine along with the OWL ontology definitions to derive additional statements as depicted in Fig. 5.

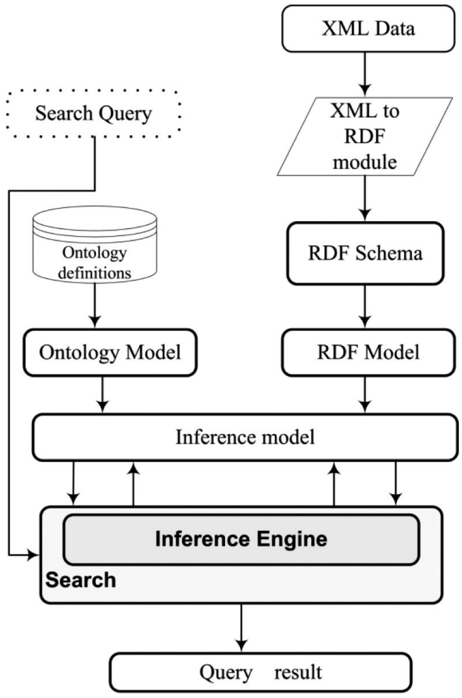


Fig. 5. SemFARM Search Module.

All statements and resources are searched for the required information about a file in query. The search is performed by navigating the inference model using Jena APIs for a specific property associated with a resource. The same process is repeated for all connected devices on which the search is intended. SemFARM supports Bluetooth connectivity to share and transfer files between connected and authorized devices. The reasoning task is performed on a network server as low-end devices are unlikely to perform reasoning in rational time because of their limited computing resources. Therefore, an XML file containing the meta-data of files is sent to the server where the reasoning is performed by binding it with ontology definitions. The search agent handles all requested queries independently and the results are automatically sent back to the requesting device. This module currently support to use single ontology but another module can be integrated to map more than one ontology or to extract information from other ontologies developed in similar domains.

M. Bhatt et al. proposed and demonstrated the prospect of sub ontology extraction from a base-ontology [29]. Similarly, A. Flahive et al. demonstrated the sub ontology extraction and extending it with new features through service oriented architecture [7] and also proposed a distributed framework for ontology tailoring [6].

5.2.1. Matching degree in SemFARM

The matching degree can be computed between properties assigned to files with the properties used in file retrieval queries based on ontology definitions. Let

- p_Q is a property used in a file retrieval query
- p_A is a property associated with a file.

The following relationships between p_Q and p_A are based on the work proposed by Paolucci et al. [26].

– *Exact match:* p_Q and p_A are equivalent, or p_Q is a subclass of p_A.
– *Plug-in match:* p_A subsumes p_Q.
– *Subsume match:* p_Q subsumes p_A.
– *Nomatch:* There is no subsumption between p_Q and p_A.

Li et al. [24] further defined match degrees by considering the semantic distance between properties in an advertisement and query, which they used for service discovery to quantify the relationships. Similarly, we can also quantify the match degrees between a property associated with a file and the properties used in a file retrieval query. For this purpose, a numerical degree is assigned for each match to quantify the relationship between p_Q and p_A. To consider the semantic distance between p_Q and p_A in assigning a match degree, Let

– *dom*(p_Q, p_A) be the degree of a math between p_Q and p_A and
– $||P_Q, P_A||$ be the semantic distance between p_Q and p_A in terms of domain ontology Ω.

Following the proposed work in [24], *dom*(p_Q, p_A) is defined for a match degree calculation as follows:

$$dom(P_Q, P_A) = \begin{cases} 1 & exactmatch, \\ \frac{1}{2} + \frac{1}{e^{(\|PQ,PA\|-1)}} & pluginmatch, \|PQ, PA\| \geqslant 2, \\ \frac{1}{2\times e^{(\|PQ,PA\|-1)}} & subsumematch, \|PQ, PA\| \geqslant 1, \\ 0.5 & uncertainmatch, \\ 0 & nomatch. \end{cases} \quad (1)$$

According to (1), for a plug-in match between p_Q and p_A, *dom*(p_Q, p_A) (0.5, 1). For a subsume match between p_Q and p_A, *dom*(p_Q, p_A) (0, 0.5).

6. Performance evaluation

Various tests and comparisons are outlined in the following subsections to measure the efficiency of SemFARM in terms of file retrieval and search accuracy. Generally, it is expected that file retrieval will be more convenient in terms of effort and time, using a keyword based search compared to manually browsing all directories.

6.1. Computing matching degree

The match degrees used for relationships between properties p_Q and p_A to retrieve a file in SemFARM can be *exact*, *plug-in* or *subsume* which are described in Section 5.2.1.

Figure 6 shows the ontology definitions used in this case study describing the classifications of health, entertainment, academic, event, health properties and personal properties fragments.

Each file on a device is annotated with two keywords. To evaluate the match degree we performed two groups of tests; namely *set 1* and *set 2*. In the first group of tests all queries were relevant to health related files to find the matching degree in one fragment, while in the second group of tests, queries were related to the personal related fragment. The queries selected for the first group includes the keywords "*treatment*", "*gp*", "*hospital*", " *tablets*" and "*health*". It can be seen in Fig. 6 that properties having an exact match to these queries includes *c3, c4, c2, c5* and $c1$. The match degree can be calculated between the associated keywords and query keywords using equation (1) described in Section 5.2.1 and in Fig. 6. Table 1 shows the matching degree calculated for four returned files as an example.

Table 1
Match degrees calculations for test set-1

Properties	c3	c4	c2	c5	c1
Files names					
file10	87%(P)	18%(S)	50%(S)	100%(E)	50%(S)
file11	50%(S)	87%(P)	100%(E)	18%(S)	87%(P)
file12	87%(P)	100%(E)	50%(S)	18%(S)	50%(S)
file13	100%(E)	87%(P)	18%(S)	50%(S)	87%(P)

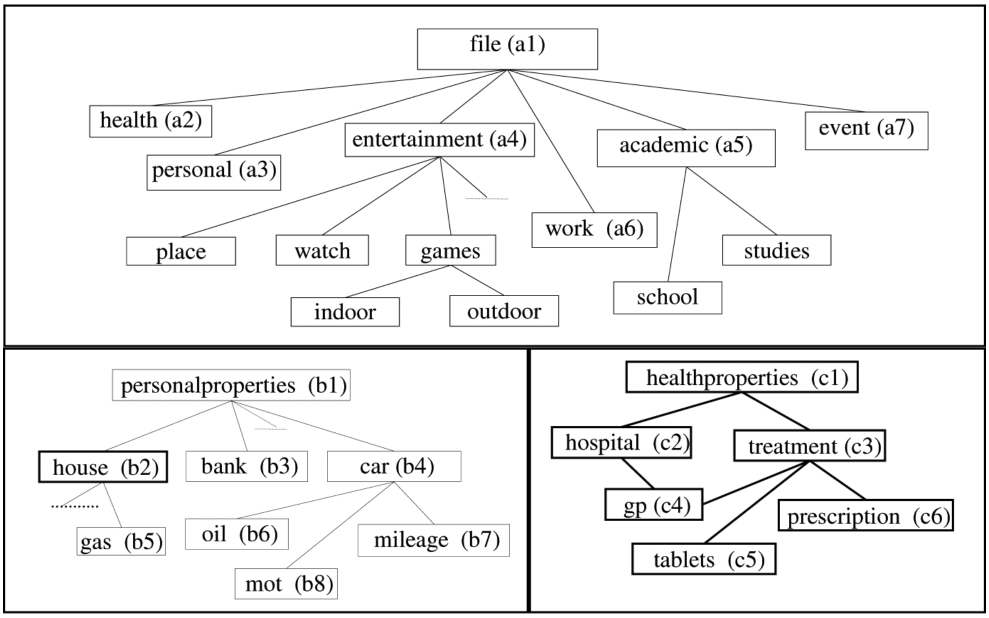


Fig. 6. Ontologies used in SemFARM.

Similarly, in the second group of tests, the selected keywords as a query were, *"mileage"*, *"car"*, *"bank"*, *"house"* and *"personal"*. Figure 6 shows the matching properties which include *b4, b7, b3, b2* and *b*1. The matching degree calculated matching degree in this fragment of personal related keywords can be seen in Table 2.

6.2. Calculating precision and recall

Precision and recall are widely used in information retrieval to evaluate the accuracy of a search mechanism [11,12,43]. To evaluate the precision and recall in SemFARM, we randomly selected 15 files and annotated them with relevant keywords which were not necessarily defined by our ontology. After executing a search query, the list of returned files was checked for relevant files and the number of relevant files was noted. The process was repeated, varying dissimilar search queries to ensure a different number of returned files for calculating the recall value from 0.1 to 1. The same test was repeated for 10 and 12 randomly selected files and the final precision and recall for SemFARM was calculated by the mean values of all three tests.

The precision and recall for untagged systems was calculated by formulating tests in which two sets of filenames, namely target-set and retrieved-set, were randomly selected. The target-set had 10 filenames while the number of files in the retrieved-set varied from 1 to 10. For each retrieved-set, the relevance

Table 2
Match degrees calculations for test set-2

Properties	b4	b7	b3	b2	b1
Files names					
file1	100%(E)	87%(P)	18%(S)	18%(S)	50%(S)
file2	50%(S)	100%(E)	6%(S)	6%(S)	50%(S)
file3	18%(S)	18%(S)	100%(E)	50%(S)	87%(P)
file4	18%(S)	6%(S)	18%(S)	100%(E)	87%(P)
file5	87%(P)	64%(S)	87%(P)	87%(P)	100%(E)

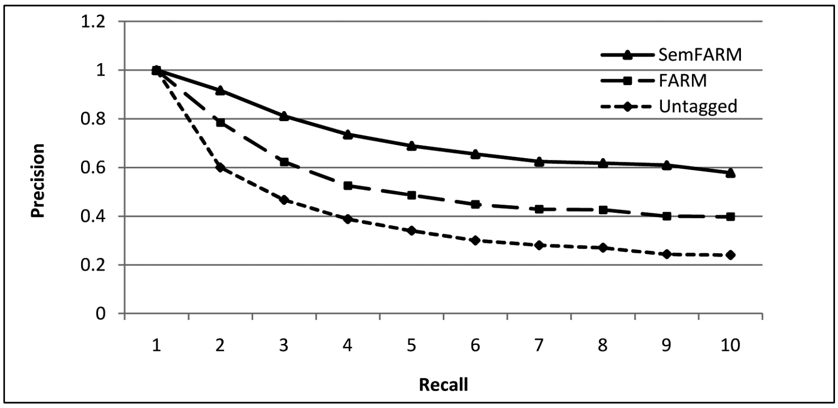


Fig. 7. Precision and Recall for Untagged, FARM and SemFARM.

was checked by comparing both sets of files. At least one file was kept relevant for all groups of tests to make sure the number of retrieved files would not be zero. The test process was implemented through a Java program and an average was taken for the 1,000 tests. For calculating precision and recall for FARM, we slightly changed the process used for calculation in the untagged system. The number of comparison of both sets was extended to three times, i.e. after the first comparison of both sets, another retrieve-set was picked up and compared with the same target-set and then a third set was picked up and compared. As discussed in Section 4.1, the annotation process annotates each file with 5 tags, in which 3 of them are annotated automatically. For this reason we compared the target-set three times with the retrieve-set for FARM as it has at least three more chances to retrieve a relevant file than the untagged system.

Let $Retrel$ be the number of relevant files retrieved, Rel be the total number of relevant files and Ret be the number of retrieved files, then recall (Rc) and $PREc$can be calculated as:

$$Rc = \frac{|Retrel|}{|Rel|}, \quad PREc = \frac{|Retrel|}{|Ret|}$$

All results are plotted in Fig. 7, which shows that precision is 1 for all three systems at 10% of recall; however, the precision is higher at most values of recalls for SemFARM followed by FARM and the untagged systems. For example, the precision of SemFARM, FARM and the untagged systems are 0.81, 0.62 and 0.46 respectively at the recall of 30%.

6.3. Probabilistic evaluation

A generalized comparison is carried out by computing the probability of a successful file search for the following approaches:

Table 3
Geometric distribution

x	1	2	3	4	5
$g_u\ (x_u; p_u)$	0.52	0.2496	0.1198	0.0575	0.0276
$g_f\ (x_f; p_f)$	0.82	0.1476	0.0265	0.0047	0.0008
$g_{sf}\ (x_{sf}; p_{sf})$	0.88	0.1056	0.01267	0.00152	0.00018

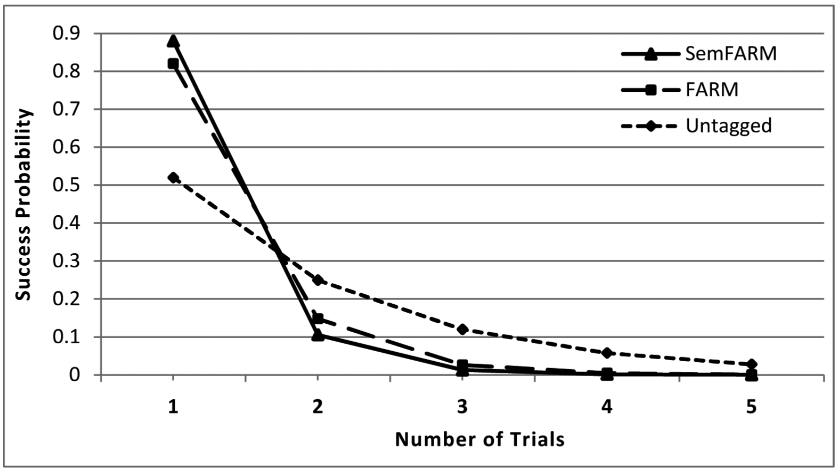


Fig. 8. Success probability of trials.

(i) *SemFAR:* semantic-based search is performed using ontology definitions on metadata which consists of file attributes and keywords.
(ii) *FARM*: search is performed on metadata which consists of file attributes and keywords.
(iii) *Untagged*: search is performed on metadata consists of filenames only.

To compare the number of searches a user has to make in order to get the desired file let p and q as the probability of success and failure for n independent trials, the distribution for the number of trials until the first success occurs is given by

$$g(x; p) = pqx - 1 \forall x = 1, 2, 3 \ldots \tag{2}$$

where g is geometric distributed variable.

If p_{sf} and q_{sf} are the success and failure probabilities, then g_{sf} is the distribution for the probability of x_{sf}^{th} trial being the first successful search for SemFARM. Similarly, for (2) the notations $\{p_f, q_f, x_f, g_f\}$ and $\{p_{u,} q_u, x_u, g_u\}$ are used for FARM and untagged systems respectively. In order to find the success probability, 100 file search trials were carried out for each of the three systems in which SemFARM, FARM and untagged system returned 88, 82 and 52 queries successfully. Table 3 presents the geometric distribution calculated for first 5 trials of each system.

It is evident from Fig. 8 that the success probability of SemFARM is higher when the number of trials is less i.e. the area under the curve is more during the first three trials as compared to the area under curve for FARM and untagged systems. This indicates that the probability of success for SemFARM is higher for lesser number of trials as compared to the rest of two systems. In other words, the chances for getting a successful query in *lesser number of trials* are greater for SemFARM when compared to FARM and untagged approaches.

The geometric distribution presented in Table 3 is based on success probability p, which can be used to symbolize the general perception about the performance of SemFARM by calculating its maximum likelihood estimation.

6.3.1. Parameter estimation for geometric distribution

The maximum likelihood estimator of p for geometric distribution is based on results presented in Table 3. For a random sample $x_1, x_2, x_3 \ldots x_n$ from a geometric distribution, the likelihood function is given by

$$\begin{aligned} L(p) &= (1-p)^{x_1-1}\, p\, (1-p)^{x_2-1} \ldots . (1-p)^{x_n-1} p \\ &= p^n (1-p)^{\sum x_i - n}, \ (0 \leqslant p \leqslant 1) \end{aligned}$$

Taking the natural logarithm $L(\theta)$

$$= \ln L(p) = n \ln p + \left(\sum_{i=1}^{n} x_i - n \right) \ \ln \ (1-p) \ , \ (0 \leqslant p \leqslant 1)$$

After taking the derivative with respect to p

$$\frac{d \ln L(p)}{dp} = \frac{n}{p} = \frac{\sum_{i=1}^{n} x_i - n}{1-p} = 0$$

After solving for p, we get

$$p = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{x}$$

The maximum likelihood estimator of p is

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

$\hat{p} = 1/x$ and therefore,

$$x = 1/\hat{p} \tag{3}$$

Using Eq. (3), we can calculate the required number of trials on which the first success is expected. For example, the value of $\hat{p}$ is 0.88 for SemFARM which means the first success is expected in 1.13 trials. Similarly the first success is 1.21 and 1.92 for FARM and untagged systems respectively.

7. Conclusion

SemFARM is presented in this paper to implement a semantic based search mechanism on mobile devices with limited user interface capabilities. SemFARM has been tested and validated on the J2ME platform. The annotation process automatically collects file attributes from the file system of the device and uses them as meta-data through the implementation of kXML. The metadata is stored locally which is then used in search for the retrieval of a required file. The MIDlet was extended to support a J2ME Bluetooth stack for searching files on other connected devices. The results and analysis prove the effectiveness and high efficiency of SemFARM. We are currently analysing the implementation of light version of a reasoner which can be integrated in SemFARM without compromising on its efficiency. The implementation of a light version reasoner can lead us to use a complete framework without depending on any network medium, because the reasoning tasks can be performed by the device itself instead of sending information to a server. We are also in the process of scrutinizing the ontology used in SemFARM to add more classes to broaden the keywords range and their definitions.

References

[1] A. Karypidis and S. Lalis, Automated context aggregation and file annotation for PAN-based computing, *Personal and Ubiquitous Computing* **11**(1) (2007), 33–44.
[2] A.P. Korpipää and J. Mäntyjärvi, An Ontology for Mobile Device Sensor- Based Context Awareness, *Modeling and Using Context: Proc. 4th Int'land Interdisciplinary Conf.* (*Context 2003*), LNCS 2680, Springer-Verlag, 2003, pp. 451–458.
[3] M. Abrams, C. Phanouriou, A.L. Batongbacal, S.M. Williams and J.E. Shuster, UIML: An Appliance- Independent XML User Interface Language, *Computer Networks* **31**(11) (1999), 1695–1708.
[4] A.B. Waluyo, B. Srinivasan and D. Taniar, Research in mobile database query optimization and processing, *Mobile Information Systems* **1**(4) (2005), 225–252.
[5] A.B. Waluyo, B. Srinivasan and D. Taniar, Research on location-dependent queries in mobile databases, *International Journal of Computer Systems: Science and Engineering* **20**(2) (2005).
[6] A. Flahive, W. Rahayu, D. Taniar and B.O. Apduhan, A Distributed Ontology Framework in the Semantic Grid Environment, Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA 2005), *IEEE Computer Society* (2005), 193–196.
[7] A. Flahive, D. Taniar, W. Rahayu and B.O. Apduhan, Ontology tailoring in the Semantic Grid, *Computer Standards & Interfaces* **31**(5) (2009), 870–885.
[8] A. Wilhelm, Y. Takhteyev, R. Sarvas, N. Van House and M. Davis, *Photo Annotation on a Camera Phone*, CHI '04 extended abstracts on Human factors in computing systems, Vienna, Austria, 2004.
[9] A. Santangelo, A. Gentile, G. Vella, N. Ingraffia and M. Liotta, XPL the eXtensible presentation language, *Mobile Information Systems* **5**(2) (2009), 125–139.
[10] B. McBride "Jena", *IEEE Internet Computing*, July/August, 2002.
[11] C. van Rijsbergen, *Information Retrieval*, Butterworths, 1979.
[12] D.A. Buell and D.H. Kraft, *Performance Measurement in a Fuzzy Retrieval Environment*, Proc. ACM SIGIR '81, 1981, pp. 56–62.
[13] Extensible Markup Language, http://www.w3.org/XML/, Accessed January 2011.
[14] F.D. Ontologies, *Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
[15] Flickr Photo Sharing, http://www.flickr.com.
[16] B. Guo, S. Satake and M. Imai, Home-explorer: ontology-based physical artifact search and hidden object detection system, *Mobile Inf Syst* **4**(2) (2008), 81–103.
[17] http://www.w3.org/TR/owl-features/ Accessed January 2011.
[18] S. Izumi, K. Yamanaka, Y. Tokairin, H. Takahashi, T. Suganuma and N. Shiratori, Ubiquitous Supervisory System based on Social Contexts using Ontology, *Mobile Information Systems* (*MIS*) **5**(2) (2009), 141–163.
[19] J.J. Carroll et al., Jena: *Implementing the Semantic Web Recommendations*, tech. report HPL-2003-146, Hewlett Packard Laboratories Bristol, 2003.
[20] S. Jan, Li. Maozhen, G. Al-Sultany and H. Al-Raweshidy, File annotation and sharing on low-end mobile devices, Fuzzy Systems and Knowledge Discovery (FSKD), *Seventh International Conference on* **6** (10–12 Aug 2010), 2973–2977.
[21] Jena2. http://www.hpl.hp.com/semweb/jena2.htm, Accessed January 2011.
[22] Jena inference: http://jena.sourceforge.net/inference/ Accessed January 2011.
[23] kXML, http://kxmlrpc.objectweb.org/, Accessed January 2011.
[24] M. Li, B. Yu, O. Rana and Z. Wang, Grid service discovery with rough sets, *IEEE Trans Knowl Data Eng* **20**(6) (2008), 851–862. doi:10.1109/TKDE.2007.190744.
[25] M.C. Daconta, L.J. Obrst and K.T. Smith, *The Semantic Web: a guide to the future of XML, Web services, and knowledge management*, Indianapolis, Ind.: Wiley Pub, 2003.

[26] M. Paolucci, T. Kawamura, T. Payne and K. Sycara, *Semantic Matching of Web Service Capabilities*, Proc. First Int'l Semantic Web Conf. (ISWC '02), June 2002, pp. 333–347.
[27] M. Smith, C. Welty and D. McGuiness, *Web Ontology Language* (*OWL*) *Guide*, August 2003.
[28] D.L. McGuinness, R. Fikes, J. Hendler and L.A. Stein, DAML+OIL: An Ontology Language for the Semantic Web, *IEEE Intelligent Systems* **17**(5) (2002), 72–80.
[29] M. Bhatt, A. Flahive, C. Wouters, W. Rahayu and D. Taniar, MOVE: A Distributed Framework for Materialized Ontology View Extraction, *Algorithmica* **45**(3) (2006), 457–481.
[30] M. Kobayashi and K. Takeda, Information retrieval on the web, *ACM Computing Surveys* (*CSUR*) **32**(2) (June 2000), 144–173.
[31] F. Monaghan and D. O'Sullivan, *Automating Photo Annotation using Services and Ontologies*, Proceedings of Mobile Services and Ontologies Workshop, 2006.
[32] M. Ames and M. Naaman, *Why we tag: motivations for annotation in mobile and online media*, Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, California, USA, 2007.
[33] N. Weißenberg, A. Voisard and R. Gartmann, Using ontologies in personalized mobile applications, in: *Proc of the Intl ACM GIS Symposium*, D. Pfoser and I. Cruz, eds, ACM Press: New York, 2004.
[34] M. Naaman and R. Nair, ZoneTag's collaborative tag suggestions: what is this person doing in my phone? *IEEE Multimed* **15**(3) (2008), 34–40.
[35] N.N. Qadri, M. Altaf, M. Fleury and M. Ghanbari, Robust video communication over an urban VANET, *Mobile Information Systems* **6**(3) (2010), 259–280.
[36] Parsing XML in J2ME, by Jonathan Knudsen, http://developers.sun.com/techtopics/mobility/midp/articles/parsingxml/ Accessed January 2011.
[37] Prashanth Mohan, Raghuraman, S. Venkateswaran and Arul Siromoney, *Semantic File Retrieval in File Systems Using Virtual Directories*, in the *Poster Session* of the 13th Annual IEEE International Conference on High Performance Computing (HiPC), Bangalore, India, Dec 2006.
[38] Resource Description Framework (RDF), http://www.w3.org/RDF/. Accessed July 2010.
[39] R.A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1999.
[40] S. Caballé, F. Xhafa and L. Barolli, Using mobile devices to support online collaborative learning, *Mobile Information Systems* **6**(1) (2010), 27–47.
[41] S. Izumi, K. Yamanaka, Y. Tokairin, H. Takahashi, T. Suganuma and N. Shiratori, Ubiquitous supervisory system based on social contexts using ontology, *Mobile Information Systems* **5**(2) (2009), 141–163.
[42] S. Schenk, O. G¨orlitz and S. Staab, *TagFS: Bringing Semantic Metadata to the Filesystem*, In Poster at the 3rd European Semantic Web Conference (ESWC), 2006.
[43] L. SU, The relevance of recall and precision in user evaluation, *Journal of the American Society for Information Science* **45** (1994), 207–217.
[44] T. Iwamoto, G. Suzuki, S. Aoki, N. Kohtake, K. Takashio and H. Tokuda, *uPhoto: A Design and Implementation of a Snapshot Based Method for Capturing Contextual Information*, presented at The Second International Conference on Pervasive Computing, Advances in Pervasive Computing, Linz/Vienna, Austria, 2004.
[45] T. Berners-Lee, J. Hendler and O. Lassila, *The Semantic Web*, Scientific American, May 2001.
[46] W3C: http://www.w3.org/ Accessed January 2011.
[47] W. Viana, S. Hammiche, B. Moisuc, M. Villanova-Oliver, J. Gensel and H.Martin, *Semantic Keyword-Based Retrieval of Photos Taken with Mobile Devices*, InMoMM, 2008, pages 192–199.
[48] ZoneTag, http://zonetag.research.yahoo.com/.

Sadaqat Jan is a PhD student in the School of Engineering and Design at Brunel University, UK. He received his Master degree in Computer Systems Engineering from NWFP University of Engineering and Technology Peshawar, Pakistan. His research interests include Mobile Computing, Information retrieval, Distributed Systems and Semantic Web. He is a member of IEEE.

Maozhen Li is a Senior Lecturer in the School of Engineering and Design at Brunel University, UK. He received the PhD from Institute of Software, Chinese Academy of Sciences in 1997. His research interests are in the areas of grid computing, intelligent systems, P2P computing, semantic web, information retrieval, content based image retrieval. He has over 80 scientific publications in these areas. He has served over 30 international conferences. He is currently on editorial boards of three journals – the International Journal of Grid and High Performance Computing, the International Journal of Distributed Systems and Technologies, and International Journal on Advances in Internet Technology. He is a member of IEEE and a Fellow of the British Computer Society.

Ghaidaa Al-Sultany has joined Brunel University- UK in 2008 as a PhD student in school of Engineering and Design. She received her MSc in computer science from Babylon University, Iraq in 2002. She worked as a lecturer in the same university

for 6 years. Her research interests are in areas of mobile software agents, semantic web computing, intelligent systems and Mobile Computing.

Hamed Al-Raweshidy is a research professor and the head of the Wireless Networks & Communications Centre in the School of Engineering and Design at Brunel University.

His research interests include network optimization, Radio over Fibre, PAN and PN, Spectral Efficiency, Green Radio, Cognitive Communications, Mesh & Ad-Hoc Networks and Future Network. He has published over 250 papers in International Journals and referred conferences. He is the editor of the first book in Radio over Fibre Technologies for Mobile Communications Networks and contributed to chapters for 6 books. Currently, he is the "Editor-in-Chief" of Communication Networks Journal /USA. He is a member of several Journal Editorial Boards such as Journal of Communications and Mobile Computing and Journal of Wireless Personal Communications. He is a member of several International Conference Advisory Committees and Technical Program Committees such as VTC, PIMRC, GLOBECOM, WPMC and IST. He has been involved in projects with several companies and operators such as Vodafone (UK), Ericsson (Sweden), Andrew (USA), NEC (Japan), Nokia (Finland), Siemens (Germany), Franc Telecom (France), Thales (UK & France) and Tekmar (Italy). He was the principal investigator for several EPSRC projects and European project such as MAGNET EU project (IP) 2004-2008. He's a senior member of the IEEE, an IEE fellow, and a member of New York Academy of Sciences.

Ibrar Shah received his B.Sc. and M.Sc. degrees in Computer Systems Engineering from NWFP University of Engineering & Technology, Peshawar, Pakistan in 2004 and 2007 respectively. Currently, he is with Wireless Networks & Communications Centre, Brunel University, West London UK working towards his PhD. His research mainly focuses on non-cooperative behaviour in Wireless Mesh Networks where he is working on routing and channel assignment problems by using game theoretic models for fairness and throughput optimization.