

## Enhanced Information Retrieval in Social Internet of Things Environment

Sungrim Kim<sup>1</sup> and Joonhee Kwon<sup>3\*</sup>

<sup>1</sup>*Department of Internet Information, Seoil University,  
28, Yongmasan-ro 90-gil, Jungnang-gu, Seoul, Korea*

<sup>2</sup>*Department of Computer Science, Kyonggi University,  
154-42 Gwanggyosan-ro, Yeongtong-ku, Suwon-si, Kyonggi-do, Korea  
[srkim@seoil.ac.kr](mailto:srkim@seoil.ac.kr), [kwonjh@kyonggi.ac.kr](mailto:kwonjh@kyonggi.ac.kr)*

### Abstract

*The Internet of Things brings promising challenges. Recently the new paradigm named Social Internet of Things has come. Social Internet of Things is one of novel approaches to integrate Internet of Things and Social Network. In Social Internet of Things environment, the amount of data is greatly increasing as it deals with much more information than ever before. This paper proposes a new information retrieval method in Social Internet of Things environment. We design a new retrieval algorithm using a levelized tiered things structure. The proposed structure is an enhanced tiered structure for supporting the social network of Internet of Thing. A scenario using our method is described. In this scenario, our method shows performance improvement of about 70 percent in the lowest level. Then several experiments are performed. The experimental results show that the proposed method's retrieval performance is better than other existing methods in Internet of Things environment.*

**Keywords:** *Internet of Things, retrieval method, tiered-things structure*

### 1. Introduction

Internet of Things (IoT) attracts great attention, and brings promising opportunities and challenges. The IoT is a convergence technology that spans devices, networks, platforms, and services. The IoT links the objects or people, then enabling anytime, anywhere connectivity for anything and not only for anyone [1]. The IoT can allow people and objects to seamlessly interact [2]. The IoT enables communication, interaction, and information sharing between people and people, people and objects, objects and objects. In the Internet environment, it is possible for the system itself to determine the situation, so that the operation can be performed autonomously without human intervention. The IoT has made it possible for devices around the world to acquire information and store it, in order to be able to use it at a later stage [3]. So far, numerous applications and services have been proposed. Research on IoT has important economic and social values for developing the next generation of information, network, and communication technologies.

In the IOT environment, the amount of data is greatly increasing as it deals with much more information than ever before. Communication networks facilitate easy access of information. Meanwhile, the richness of online information also brings forth the information overload problem [4]. A user might not be able to find out all the items from a lot of information available on the internet. Users need to be able to find relevant information to satisfy their particular information needs. The need of effective information retrieval method becomes essential for easy access to relevant information [5].

---

Received (January 25, 2017), Review Result (May 11, 2017), Accepted (June 4, 2017)

\* Corresponding Author

The retrieval methods are capable of helping user to easily find information among a large amount of information by determining what is of attention to a user [6]. It is important to incorporate the IoT features into the retrieval method. However, the traditional retrieval methods have not taken into account the IoT features while making retrieval. As IoT is expanding, there have been only few researches on retrieval method in IoT. Therefore, another new IoT retrieval method researches are necessary.

We design a new retrieval method in IoT environment. We adopt the tiered-things structure based on social relationship among objects. We propose a leveled tiered-things structure, where the structure is an enhanced structure from [7]. Our method improves performance of top-k retrieval by adjusting the value of k according to a user's attention level. When a user finds top-k documents while paying little attention, only some top documents of them are retrieved, not all of them [8].

The remainder of the paper is organized as follows. First, in Section 2, we describe background and related works about retrieval method and IoT. In Section 3, we explain our proposed retrieval algorithm and structure. And we illustrate a music scenario using our proposed method in IoT environment. In Section 4, we evaluate the efficiency of leveled recommendation method. Finally, Section 5 will conclude the paper.

## 2. Background and Related Works

The IoT integrates a large number of technologies. The IoT is not the result of a single novel technology. It is several complementary technical developments. The IoT helps to bridge the gap between the virtual and physical world. The IoT is the internetworking of physical devices, and network connectivity that enable these objects to collect and exchange data. It refers to a world where physical objects and beings, as well as virtual data and environments, all interact with each other in the same space and time. In the IoT, each person and thing has a locatable, addressable, and readable counterpart in the Internet. The IoT will provide a paramount set of opportunities to users, manufacturers, and service providers with a wide applicability in many areas. Recently, new applications and research challenges in numerous areas of IoT are getting started [1, 9-12].

Recently there have been quite a number of research activities that investigated the potentialities of integrating social networking concepts into IoT. Social Internet of Things(SIoT) combines the IoT with the social networks [3, 13-15].

In [15], they consider social concerns as a research challenge for IoT, specifically regarding the interaction of things in social networks. It systematically investigates the key technologies related to the development of IoT and its applications, particularly from a data-centric perspective. It reviews and compares technologies including data streams, data storage models, searching, and event processing technologies, which plays a vital role in enabling the vision of IoT.

In [16], they investigate on the potential of combining social and technical networks to collaboratively provide services to both human users and technical systems. The common physical and social space emerges by the objects' ability to interconnect, not only amongst themselves, but also with the human beings living and working in them. It introduces the CognitiveOffice as an example environment and highlight the potential of socio-technical networks.

Social networks essentially consist of a representation of each user, his/her social links, and a variety of additional services. The SIoT has the potential to support novel applications and networking services for the IoT in more effective and efficient ways. SIoT has social relationships between human beings and things, things and things, things and their owners. In SIoT, there are five social relationships are defined as follows [13],[14].

- Parental object relationship (POR): established among objects belonging to the same production batch

- Co-location object relationship (C-LOR): established among objects used always in the same place
- Co-work object relationship (C-WOR): established whenever objects collaborate to provide a common IoT application
- Ownership object relationship (OOR): established among heterogeneous objects which belong to the same user
- Social object relationship (SOR): established when objects come into contact, sporadically or continuously, because their owners come in touch with each other during their lives

The traditional retrieval methods concern themselves with documents, queries, and their relations to each other. Every information retrieval system is built from two main parts: the index structure and the retrieval/ranking algorithm. Tiered index based method is known as efficient top-k retrieval method. When using tiered indexes, we search for a document in the first tier. If we fail to get k results from it, search falls back to tier2, and so on [17]. Figure 1 shows the basic idea and process of tiered index.

- Create several tiers of indexes, corresponding to importance of indexing terms
- During query processing, start with the highest-tier index
- If we get k results then stop
- If we don't get k results, repeat for the next index

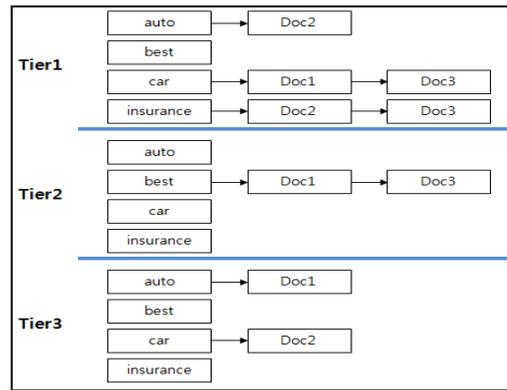
**Figure 1. The Basic Idea and Process of Tiered Index**

The tiered indexing method [6] classifies documents by dividing them based on the term frequency value priority, which is an important factor affecting the search score. In other words, when the term frequency value is high, it is classified as one of upper layers, and when the value is low, it is classified as one of lower layers. When a query term is given, only documents classified as upper layers having a higher priority are searched. And if the search results are k or more, the search is terminated and documents are classified as lower layers.

**Table 1. Example: Table of tf Values**

	Doc1	Doc2	Doc3
car	27	4	24
auto	3	33	0
insurance	0	33	29
best	14	0	17

[6] illustrates example of tiered-index using tf values (Table 1) in Figure 2. In this example we set a tf threshold of 20 for tier 1 and 10 for tier 2, meaning that the tier 1 index only has postings entries with tf values exceeding 20, while the tier 2 index only has postings entries with tf values exceeding 10. In this example we have chosen to order the postings entries within a tier by document ID. If we fail to get K results from tier 1, query processing “falls back” to tier 2, and so on. Within each tier, postings are ordered by document ID.



**Figure 2. Example: Tiered Index**

[7] proposed efficient Top-k retrieval method based on SIoT. In [7], search scores are calculated using social relations between objects. Recently, an efficient Top-k retrieval method based on SIoT was proposed [3]. It is a new retrieval method that applies the tiered indexing method to the SIoT environment. The tiered indexing method is known as one of efficient top-k information retrieval methods in traditional information retrieval. In [7], SIoTScore, which is one of search score formulas, is proposed. An important factor affecting the SIoTScore value is the social relationship value SR between objects. SR is the social relationship value between things [18].

### 3. Retrieval Method in Internet of Things

#### 3.1. Structure and Algorithm

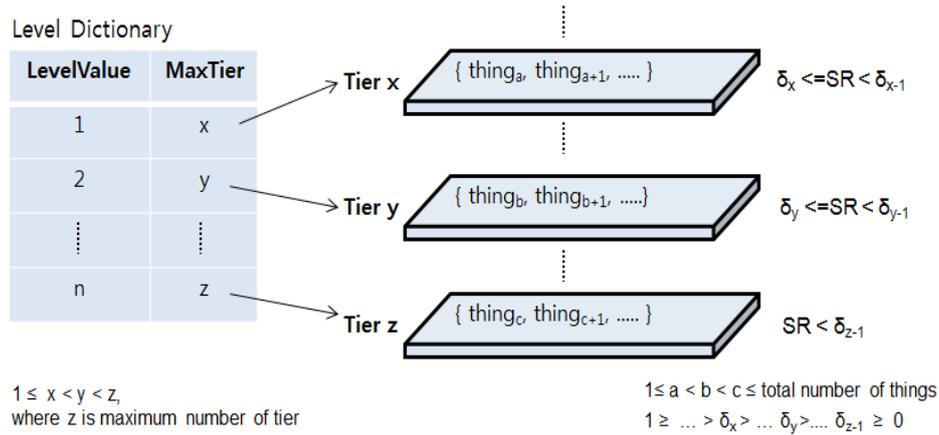
This section presents an enhanced information retrieval method in SIoT environment. We adopt the tiered things structure of [7], where the structure is based on social relationship between objects.

Our method improves performance of top-k retrieval by adjusting the value of k according to a user's attention level. When a user finds top-k documents while paying little attention, only some top documents of them are retrieved, not all of them.

Our adjusted value of k is denoted by  $k(i)$ . In Equation (1),  $k(i)$  is the levelized k value, when current level value is i and the maximum number of level values is l. In our levelizing scheme, the degree of attention is denoted by a level value. The higher the attention is, the higher the level value is.

$$k(i) = k/l * i \tag{1}$$

We propose a levelized tiered things structure, where the structure is an enhanced structure from [7]. Figure 3 shows it. The tiered things on the right of this figure are them in [7], where they are based on social relation value, SR. In this paper, a level dictionary is proposed. The level dictionary is composed of 'LevelValue' and 'MaxTier', where the former is current level value and the latter is maximum tier number for this level value.



**Figure 3. Levelized Tiered Things Structure**

Table 2 shows our retrieval algorithm using the levelized tiered things structure. When a user is in level ‘L’, our method finds documents owned by things in tier 1. When the size of the result set is less than  $k(L)$ , query processing falls back to tier 2, and maximum tier number for level value L in level dictionary.

**Table 2. Retrieval Algorithm Using Levelized Tiered Things Structure**

```

Algorithm.
Begin
Input
    Q : query
    Tiered-Things : tiered things
    k : the number of result to be retrieved
    LD : Level Dictionary
    L : number of level
Output
    ResultSet : top-k result document set
Method
    tier = 1;      MaxLevel = maximum value of LevelValue in LD;
    ResultSet = { };
    while (tier ≤ LD[L].MaxTier)
    {
        for (each j ∈ Tiered-Thingsx)
            ResultSet = ResultSet ∪ resultSetj(Q);
        if ( |ResultSet| < k/MaxLevel*L )
            tier ++;
        else
            break;
    }
    return topK(ResultSet);
End.
    
```

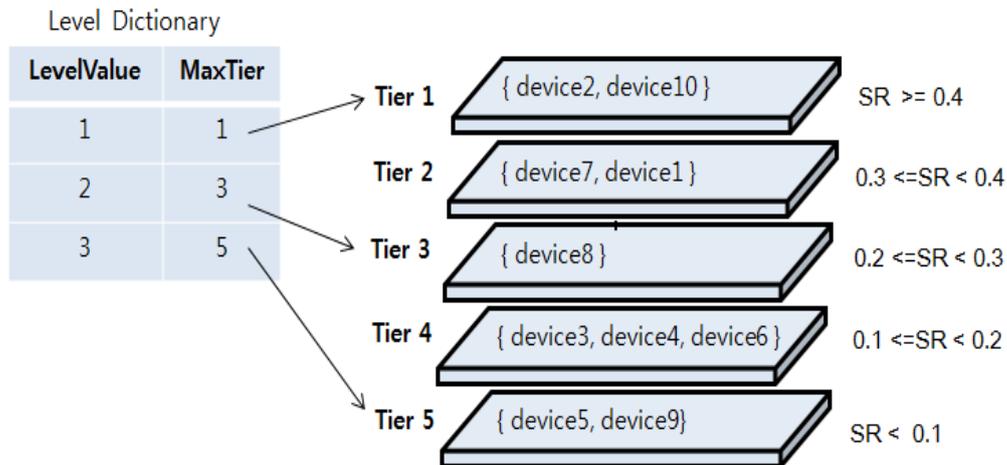
**3.2. Scenario**

This section presents Jane’s music scenario using our proposed method. Table 3 shows her friend’s device information, SR value of devices and documents in them.

**Table 3. Device Information**

Device	SR	Documents in device
device1	0.32	doc1
device2	0.44	doc2
device3	0.18	doc3
device4	0.11	doc4
device5	0.05	doc5, doc6
device6	0.17	doc7, doc8
device7	0.34	doc9, doc10
device8	0.28	doc11
device9	0.03	doc12
device10	0.60	doc13, doc14, doc15

We suppose the following levelizing policy. Just finding related top-k documents using search query is supposed to be in level 1. Then, searching for query while listening to the query related music is supposed to be in level 2, because she pays more attention in the query than just searching for it. Next, searching for query while she listens to the query related music for a long time is supposed to be in level 3. Figure 4 shows the levelized tiered things structure in this scenario.



**Figure 4. Levelized Tiered Things Structure Based on SR in Table 2**

In this scenario, we suppose that we find related top-9 documents for query ‘Michael Jackson’. Table 4 shows the document information for it, that is, SR value of device owned the document, document score with respect to the query, and SIoTScore. To retrieve top-9 documents using traditional retrieval method in this scenario, 15 documents are accessed.

**Table 4. Document Information for Query ‘Michael Jackson’**

Document	Device	SR	score	SIoTScore
doc1	device1	0.32	0.44	0.14
doc2	device2	0.44	0.95	0.41
doc3	device3	0.18	0.61	0.11
doc4	device4	0.11	0.07	0.01
doc5	device5	0.10	0.17	0.02
doc6	device5	0.10	0.04	0.00
doc7	device6	0.17	1.00	0.17
doc8	device6	0.17	0.23	0.04
doc9	device7	0.34	0.09	0.03
doc10	device7	0.34	0.03	0.01
doc11	device8	0.28	0.87	0.24
doc12	device9	0.03	0.60	0.02
doc13	device10	0.60	0.85	0.51
doc14	device10	0.60	0.32	0.19
doc15	device10	0.60	0.56	0.34

First, Jane just finds related top-9 documents using query ‘Michael Jackson’, so level value becomes 1. In level 1, at most, only top-3 documents are retrieved, according to equation (1). As shown in Table 3, only 4 documents in device2 and device10 are accessed, using levelized tiered things structure in Figure 4. In this level, top-3 documents are retrieved from the 4 documents accessed in tier1, so results are doc13, doc2, and doc 15.

**Table 5. Documents Accessed for Top-9 Retrieval (Level 1)**

Tier	Device	Document	SR	score	SIoTScore
Tier1	Device2	Doc2	0.44	0.95	0.41
	Device10	Doc13	0.60	0.85	0.51
	Device10	Doc14	0.60	0.32	0.19
	Device10	Doc15	0.60	0.56	0.34

Secondly, she searches for the query ‘Michael Jackson’ while listening to his music ‘Billie Jean’. Now, the level value becomes 2. In the level, the number of top documents retrieved is increased, from 3 to 6, because she pays more than the attention in level 1. Table 6 shows the documents accessed for retrieval in level 2. In level 2, only documents in tier 1 and tier 2 are accessed because only top-6 documents are required in level 2.

Documents retrieved in level 2 are doc13, doc2, doc15, doc14, doc1, doc9. The number of documents accessed in this level is 7.

**Table 6. Documents Accessed for Top-9 Retrieval (Level 2)**

Tier	Device	Document	SR	score	SIoTScore
Tier1	Device2	Doc2	0.44	0.95	0.41
	Device10	Doc13	0.60	0.85	0.51
	Device10	Doc14	0.60	0.32	0.19
	Device10	Doc15	0.60	0.56	0.34
Tier2	Device1	Doc1	0.32	0.44	0.14
	Device7	Doc9	0.34	0.09	0.03
	Device7	Doc10	0.34	0.03	0.01

Thirdly, she searches for ‘Michael Jackson’, while she continues to listen to his music ‘Beat It’ after listening to ‘Billie Jean’. So, the level value becomes 3. In this level, top-9 documents are retrieved like traditional top-k retrieval. Documents in tier 5 are not accessed because only top-9 documents are required, so the number of documents accessed is 12. Documents retrieved in level 3 are doc13, doc2, doc15, doc11, doc14, doc1, doc7, doc3, doc8.

**Table 7. Documents Accessed for Top-9 Retrieval (Level 3)**

Tier	Device	Document	SR	score	SIoTScore
Tier1	Device2	Doc2	0.44	0.95	0.41
	Device10	Doc13	0.60	0.85	0.51
	Device10	Doc14	0.60	0.32	0.19
	Device10	Doc15	0.60	0.56	0.34
Tier2	Device1	Doc1	0.32	0.44	0.14
	Device7	Doc9	0.34	0.09	0.03
	Device7	Doc10	0.34	0.03	0.01
Tier3	Device8	Doc11	0.28	0.87	0.24
Tier4	Device3	Doc3	0.18	0.61	0.11
	Device4	Doc4	0.11	0.07	0.01
	Device6	Doc7	0.17	0.75	0.13
	Device6	Doc8	0.17	0.23	0.04

In the third level, our method is exactly the same with the previous research [7]. However, as compared with it, our method accesses 4 documents in level 1 and 7 documents in level 2. It is a decrease of about 70 percent and 40 percent.

## 4. Experiments

### 4.1. Experimental Setup

In this section, we perform experimental evaluation. We implemented three methods, the existing retrieval method without tier, the existing method with tier in [7], and our method. All programs were written in Java.

We used 7 synthetic document sets in [7]. The value of SR and score for each document was generated randomly. The number of tiers and levels was set as 4. Figure 5 shows the levelized tiered things structure used in this experiment.

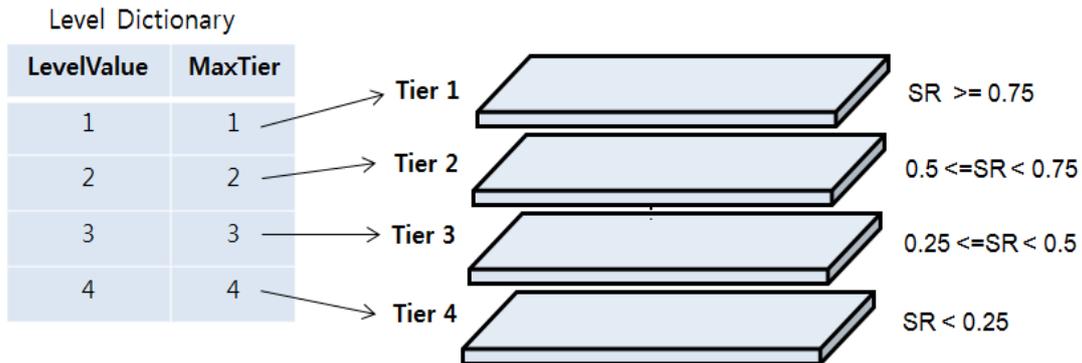


Figure 5. Levelized Tiered Things Structure in the Experimental Setup

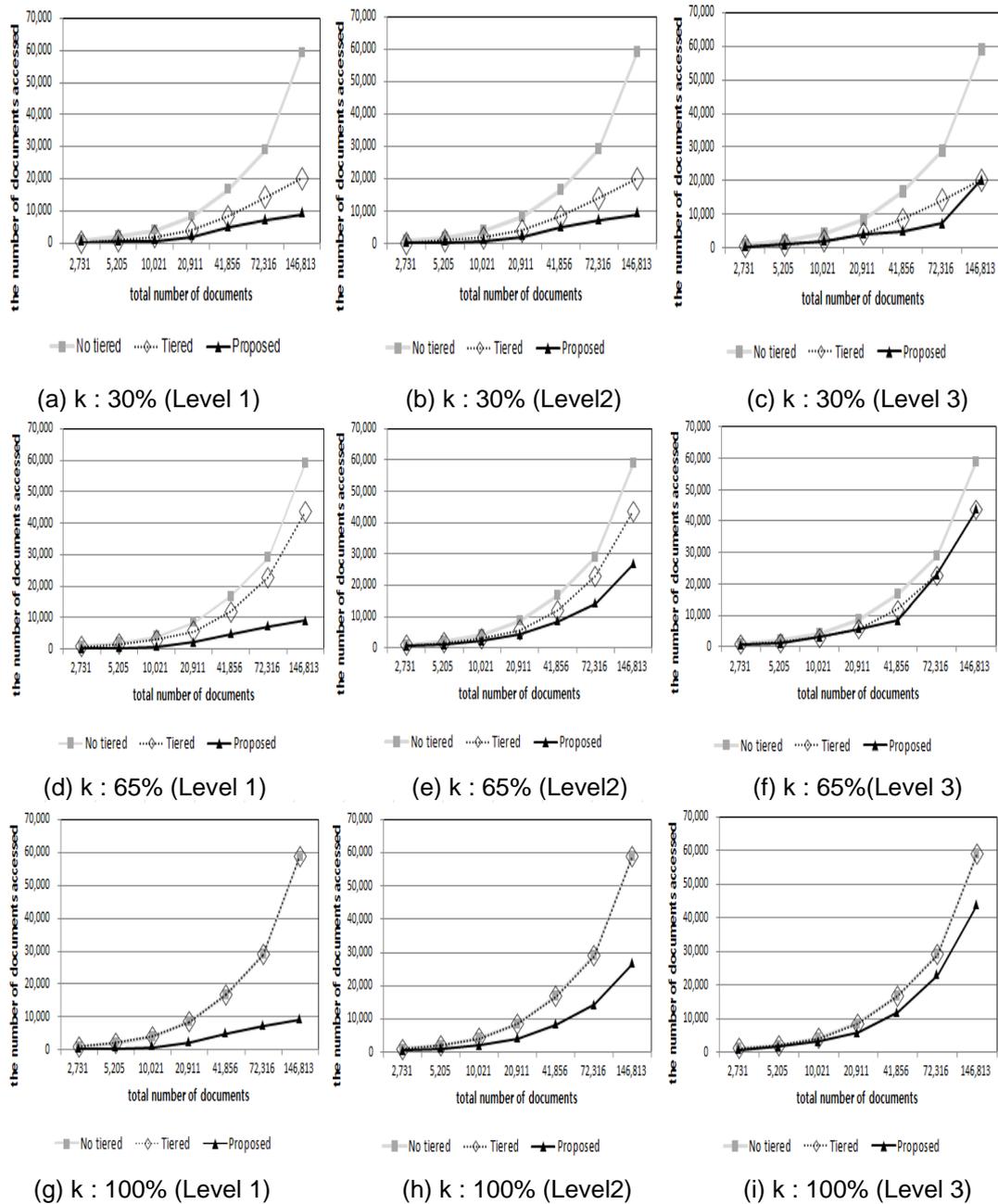
### 4.2. Experimental Results

We evaluated the rapidness through comparative experiments. We measured the number of documents accessed whenever the number of documents is increased. Figure 6 shows the experimental results. In it, the existing method without tier is called 'No tiered', the existing method with tier is called 'Tiered', and the proposed method is called 'Proposed'.

For comparison, we varied the value of  $k$  of top- $k$  retrieval, where  $k$  was varied by 30 percentage, 65 percentage, and 100 percentage of total number of documents. When the level is in 4, both 'Tiered' and 'Proposed' method show the same result. That is the reason why results in level 4 are not depicted in Figure 6.

The experimental results can be indicated as follows. First, the number of documents accessed of the proposed method is consistently smaller than it of the existing methods. Second, as the total number of documentation is increased, the result curve of our method rises more gently than it of existing methods. The most noteworthy point is that the difference between the result of the proposed method and it of the existing methods increases as the level value is low. Third, we observe a result curve of 'Tiered' rises rapidly when the value of  $k$  increases. The 'Tiered' method does not show the good result when value of  $k$  is high.

It is generally acceptable that 'Tiered' method shows outstanding performance than existing 'No tiered' method, but it is not always established in high- $k$  value. Compared with it, our method is superior to other methods even though the value of  $k$  is high.



**Figure 6. The Number of Items Accessed**

## 5. Conclusion

The IoT is a new technology paradigm and is recognized as one of the most important areas of future technology. The basic idea of IoT is the pervasive presence around us of a variety of things or objects. Retrieval method is the activity of obtaining resources relevant to an information need from a collection of resources. The IoT is a new promising technology made from a variety of technology, bringing changes in retrieval method. However, the existing retrieval methods do not take IoT characteristics into account and do not guarantee good results in IoT environment.

This paper presents the new retrieval method in IoT environment. Firstly, we propose a leveled tiered-things structure based on the social relationship among objects. The social relationship is one of the most important IOT features. The tiered-things structure is one of the efficient top-k retrieval methods. In the proposed method, the social relationship

and the tiered-things structure are integrated so that the user is efficiently provided with relevant information. Secondly, we propose retrieval algorithm using the levelized tiered things structure. When users find top-k results, our algorithm adjusts the value of k according to a user's attention level. When a user finds top-k results while paying little attention, only some of more important and relevant results are retrieved, not all of them. It improves performance of top-k retrieval in IoT environment. Thirdly, we explained our method with a music scenario. This scenario showed performance improvement of about 70 percent in the lowest level. Finally, several experiments were performed and the experimental results showed that our method had a good performance.

## Acknowledgments

This paper is a revised and expanded version of a paper entitled [A Study on Retrieval Method in Internet of Things] presented at [SIT 2016, Jeju Island, Korea, December 21-23, 2016]. The present research has been conducted by the Research Grant of Seoil University in 2017.

## References

- [1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey", *Computer Networks*, Vol. 54, No. 15, (2010), pp. 2787-2805.
- [2] F. Zhao, Z. Sun, and H. Jin, "Topic-centric and semantic-aware retrieval system for internet of things", *Information Fusion*, vol. 23, (2015), pp. 33-42.
- [3] D. Gil, A. Ferrández, H. Mora-Mora, and J. Peral, "Internet of Things: A Review of Surveys Based on Context Aware Intelligent Services", Published online 2016 Jul 11. doi:10.3390/s16071069, vol. 16, no. 7, (2016), pp. 1069.
- [4] X. Yang, Y. Guo, Y. Liu and H. Steck, "A survey of collaborative filtering based social recommender systems", *Computer Communications*, vol. 41, (2014), pp. 1–10.
- [5] K. Nagwekar, "A Survey on Recommendation Systems based on Online Social Communities", vol. 4, no. 12, (2016).
- [6] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press. (2008).
- [7] J. H. Kwon and S. R. Kim, "Efficient Top-k Retrieval Method Based on Social Internet of Things", *Journal of Korean Institute of Information Technology*, vol. 14, no. 6, (2016), pp.103-110.
- [8] S. Kim and J. Kwon, "A Study on Retrieval Method in Internet of Things", *Proceeding of International Conference SIT (2016)*, pp.88-91
- [9] [https://en.wikipedia.org/wiki/Internet\\_of\\_Things](https://en.wikipedia.org/wiki/Internet_of_Things)
- [10] Luigi Atzori, Antonio Iera, and Giacomo Morabito., "Making things socialize in the Internet — Does it help our lives?", *Kaleidoscope 2011: The Fully Networked Human? - Innovations for Future Networks and Services*, Proceedings of ITU, (2011), pp.1-8.
- [11] Friedemann Mattern and Christian Floerkemeier, "From the Internet of Computers to the Internet of Things", Vol. 6462 of the series *Lecture Notes in Computer Science*, (2010), pp.242-259
- [12] Luigi Atzori, Antonio Iera, and Giacomo Morabito, "From "Smart Objects" to "Social Objects": The Next Evolutionary Step of the Internet of Things" *IEEE Communications Magazine*, (2014), pp.97-105
- [13] Luigi Atzori, Antonio Iera, Giacomo Morabito and Michele Nitti. "The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization", *Computer Networks*, (2012), pp.3594-3608.
- [14] Luigi Atzori, Antonio Iera and Giacomo Morabito. "SIoT: Giving a Social Structure to the Internet of Things", *IEEE Communications Letters*, Vol. 15, No. 11, (2011), pp.1193-1195.
- [15] Y. Qin, Q. Z. Sheng, N. J. Falkner, S. Dustdar, H. Wang and A. V. Vasilakos, "When things matter: A survey on data-centric internet of things.", *Journal of Network and Computer Applications*, (2016), pp.137–153
- [16] M. Kranz, L. Roalter, and F. Michahelles, "Things that Twitter: Social Networks and the Internet of Things", *Proc. What Can the Internet of Things Do for the Citizen Wksp., 8th International Conference Pervasive Computing*, (2010).
- [17] I. badarinza and A. Sterca, "Clustering, Tiered indexes and term proximity weighting in text-based retrieval", *Studia Universitatis Babeş-Bolyai, Informatica*, vol. 57, no. 4, (2012), pp.122-130.
- [18] S. Kim and J. Kwon, "Recommendation Technique using Social Network in Internet of Things Environment", *Journal of the Korea Society of Digital Industry and Information Management*, vol. 11, no. 1, (2015), pp. 47-57.

## Authors



**Sungrim Kim**, She received her B.S., M.S. and Ph.D. in Computer Science from Sookmyung Women's University, Korea in 1994, 1997, and 2002, respectively. She is an associate professor in Department of Internet Information, Seoil University, Korea. Her research interests are Database, Information Retrieval, Social Network, Internet of Things, and Mobile Computing.



**Joonhee Kwon**, She received her B.S., M.S. and Ph.D. in Computer Science from Sookmyung Women's University, Korea in 1992, 1994, and 2002, respectively. She worked for Ssangyong Information & Communication Corporation, Korea from 1994 to 2003. She is a Professor in Department of Computer Science, Kyonggi University, Korea. Her research interests are Database, Information Retrieval, Social Network, Internet of Things, and Mobile Computing.