Summary Extract Approach of Heterogeneous Multimedia Documents Based on Folksonomy with Incentive and Quality Assurance Mechanisms

Jin-Bo BAI\textsuperscript{1}, Hong-Bo LI\textsuperscript{3,a,*}, Xu SUN\textsuperscript{2}, Shu-Sen DONG\textsuperscript{2}, Lei-Lei KONG\textsuperscript{2}, Chang-Wei WU\textsuperscript{2}, Hong DENG\textsuperscript{2}, Feng ZHAO\textsuperscript{2} and Yan-Yan ZHANG\textsuperscript{2}

\textsuperscript{1}Economics & Management College, Zhaoqing University, Zhaoqing, Guangdong, China
\textsuperscript{2}College of Computer Science and Technology, Heilongjiang Institute of Technology, Harbin, Heilongjiang, China
\textsuperscript{3}School of Computer Science and Software, Zhaoqing University, Zhaoqing, Guangdong, China

\textsuperscript{a}islhb@126.com
*Corresponding author

Keywords: Information Management, Summary Extract, Heterogeneous Multimedia Documents, Folksonomy, Incentive and Quality Assurance Mechanisms.

Abstract. Summary extract of heterogeneous multimedia documents is an important content and requirement of information management at the Internet era. Though some summary extract approaches have been proposed in recent years, in general they can only be used in a single-type media. More important is that the existed approaches of multimedia summary extract had no satisfactory effects, and can’t be brought into practice. Aiming at these problems, an approach based on folksonomy with incentive and quality assurance mechanisms is put forward. The proven folksonomy ensures the practicality of the approach, the incentive and quality assurance mechanisms avoid the innate problems of folksonomy: cold boot and inaccurate tagging. The experiment shows that the approach is reasonable and effective.

Introduction

The development of the Internet produced an explosive growth of text, image, audio, video, and other multimedia information. However, people have to encounter the embarrassment that the abundance information can’t supply with useful knowledge. With the growing and updating of the huge amounts of multimedia information, the new information processing bottleneck is how to filter and pinpoint the most appropriate and accurate information. In such a background, the value of automatic summary extract of multimedia documents is fully revealed, and has increasingly come into notice of both domestic and international academia.

In this paper heterogeneous multimedia document includes not only a single form of text, image, video, audio, but also a combination of them. From perspectives of the basic meaning, starting point and function, heterogeneous multimedia document’s summary extract is similar to text document’s automatic
summarization. The so-called automatic summarization of a text document is a computer science technology that automatically extracts the most representative sentences of the content of a text and forms them into a concise and coherent essay [1].

In essence, the summary extract of a multimedia document is a content extract, and the automatic summarization of a text document is a text extract. Though content extract has become a hot research spot, and researchers have proposed many technologies and approaches, at present these technologies and approaches can’t be brought into practice. In order to extract content from a heterogeneous multimedia document, a novel approach must be put forward.

**Related Works**

The research of automatic summary of text document, known as automatic summarization, started from the late 1950s. In order to quickly and accurately extract useful information from scientific literature, Luhn firstly proposed the concept of automatic summarization [2]. After entering the 1990s, with the vigorous development of the Internet, automated information processing has become a problem urgently to be solved. In December 1993, the first symposium of automatic summarization was convened in Wadern, Germany [3]. In 1995, the international journal of information processing & management published a special issue of summarizing text, which marked the time had come for an automatic abstract [4].

Along with the development of the era, the automatic abstract objects steered from scientific and technological literature gradually to electronic document of Internet [5-15], and the fields steered from science and technology to news, finance, entertainment, and so on [16-18]. In order to efficiently handle huge amounts of information and meet the needs of users, the new summary extract types, such as personalized summaries, update summaries, sentiment-based summaries, are emerging [19-24]. At the same time, researchers began to explore new fields, such as music summarization [25-27], video and image summarization [28-29].

There are two main problems in the above summarization technologies. The first one is that they can’t be applied in reality, because of their effects are little and the judging standard of the summary results are very subjective. The second one is that they can only be used in single media type, and can’t be used in hybrid media type, namely heterogeneous multimedia type. The second problem leads to the application scope of these technologies are very limited and can’t meet the needs of nowadays users and society.

**Main Contribution and Concrete Steps of Our Approach**

**Main Contribution**

For the sake of applying in reality, we take advantage of improved folksonomy technology in our approach. The folksonomy concept is put forward by Thomas Vander Wal, and is increasingly improved and developed by researchers. Folksonomy, also known as collaborative tagging, social classification, social indexing, and social tagging, is a system in which users apply public tags to online items, typically to aid them in re-finding those items [30]. In general, a
folksonomy gets an integrated tag set, and every tag represents a keyword of a being described item. Though folksonomy is a proven technology, it has two main insurmountable flaws: cold boot and inaccurate tagging. Cold boot means there’s not enough tags to describe an item, and inaccurate tagging means the tags can’t accurately describe an item. Why do these phenomena will happen? We argue that folksonomy lacks of two mechanisms: incentive and quality assurance.

We had proposed a multimedia retrieval approach based on tagging technology and communities performance evaluation [31]. In essence, the approach made use of an improved folksonomy: not only owners but also community colleagues can add tags to describe items, and meanwhile incentive and quality assurance mechanisms are provided.

In this paper, the proposed approach will make use of the folksonomy with incentive and quality assurance mechanisms again. So the main contribution of our paper is that we introduce the improved folksonomy from multimedia retrieval field to a new field: summary extract of heterogeneous multimedia documents.

**Concrete Steps**

The approach in this paper includes three main steps: tagging heterogeneous multimedia documents, simplifying tag set and tag ranking according to approving ratio.

**Tagging Heterogeneous Multimedia Documents.**

This step is divided into 3 sub-steps: data modeling, business logic inserting, indicators computing. In the first sub-step, data modeling, the main purpose is introducing incentive and quality assurance mechanisms to surmount the two innate flaws of folksonomy: cold boot and inaccurate tagging, and the main means is to build an E-R (Entity Relation) data model.

In the model, the data which be used to realize the incentive and quality assurance mechanisms should be described. For example, the community member entity’s attributes are (mem-ID, name), the tagging record entity’s attributes are (doc-ID, mem-ID, tag content, time, approving number, opposing number, valid status), the document entity’s attributes are (doc-ID, name, type), the approving and opposing record entity’s attributes are (doc-ID, mem-ID, tag content, app or opp).

In the second sub-step, business logic inserting, the main purpose is inserting rules. These rules are as following:

- a member can tag on any document;
- a member can add at most $m$ tags on a document ($m$ is determined by a group of experts);
- the number of approving or opposing tag of a member is no more than $n$ ($n$ is determined by a group of experts or by some kind of automatic method);
- a member can only approve or oppose others’ tag once, and can’t evaluate the tag added by himself or herself.

In the third sub-step, indicators computing, firstly according to tag’s approving ratio $r$, the valid status of a tag will be added (if $r > 1$, then add 1; otherwise, add 0); secondly a member’s tag approving number (if and only if the tag’s valid status is
1) and opposing number (if and only if the tag’s valid status is 0) will be computed, and these numbers will be taken as the member’s performance score.

The equation of the tag’s approving ratio is

\[
r = \frac{a}{o},
\]

(1)

Where \(a\) is the tag’s approving number, and \(o\) is the tag’s opposing number.

It’s worth noting that the divisor must not be 0 when the approving ratio \(r\) is computed. In our approach, if a tag’s opposing number equals 0, simply let it be 0.5.

Because of the incentive and quality assurance mechanisms being used in the improved folksonomy, the step of tagging heterogeneous multimedia documents can avoid the cold boot and inaccurate tagging problem.

**Simplifying Tag Set.**

After the former step, due to the incentive and quality assurance mechanisms, the heterogeneous multimedia documents will be added plenty of tags. So the tag set of heterogeneous multimedia documents is not concise, namely not minimum. To surmount this difficulty, the tags whose valid status is 0 must be deleted from the current tag set.

**Tag Ranking According to Approving Ratio.**

In the tag set, the approving ratio of some tag is comparatively big, and the approving ratio of some tag is comparatively small. The number is bigger means the tag is more important and more typical to the content of a heterogeneous multimedia document, and vice versa. Due to the reading habit of the vast majority of people adopting a linear way, for the sake of efficiency, the tags will be ranked in a descending order according to their approving ratios. In addition, when the approving ratios of tags are equal to each other our approach will rank the tags according to the lexicographic order.

**Experiment**

We invited 8 boys and girls to tagging the cartoon — the monkey king: uproar in heaven, and asked them that everyone can add at most 5 tags, and approve or oppose at most 35 tags. A day later, we got the initial tag set, as shown in Table 1. In Table 1, the “SN” means “serial number”, and “No. of App. / Opp.” is an abbreviation of the phrase “number of approving/opposing”.

Through the step 2, we got another tag set. In this step, some tags whose valid status value is 0 are abandoned. The abandoned tags include “毕马瘟”, “笑天犬”, and so on. So we got the simplified tag set, as shown in Table 2. In Table 2, the “SN” is ditto.

In the step 3, we ranked tags according to their approving ratios. The final tag set is as shown in Table 3. The final tag set is a summary extract of the cartoon. In fact, we can get the content of the cartoon immediately as soon as our eyes scan several tags from the beginning of the tag set. In Table 3, the “RN” means “ranking number”.

205
Table 1. The initial tag set of the cartoon.

<table>
<thead>
<tr>
<th>SN</th>
<th>Tag</th>
<th>No. of App./Opp.</th>
<th>SN</th>
<th>Tag</th>
<th>No. of App./Opp.</th>
<th>SN</th>
<th>Tag</th>
<th>No. of App./Opp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>孙悟空</td>
<td>8/0</td>
<td>13</td>
<td>哮天犬</td>
<td>5/0</td>
<td>25</td>
<td>水帘洞</td>
<td>6/1</td>
</tr>
<tr>
<td>02</td>
<td>太白金星</td>
<td>8/0</td>
<td>14</td>
<td>龙王</td>
<td>5/0</td>
<td>26</td>
<td>猪八戒</td>
<td>4/4</td>
</tr>
<tr>
<td>03</td>
<td>唐僧</td>
<td>2/6</td>
<td>15</td>
<td>东海龙宫</td>
<td>4/0</td>
<td>27</td>
<td>大圣</td>
<td>3/4</td>
</tr>
<tr>
<td>04</td>
<td>西天取经</td>
<td>3/5</td>
<td>16</td>
<td>金箍棒</td>
<td>4/4</td>
<td>28</td>
<td>齐天大圣</td>
<td>4/4</td>
</tr>
<tr>
<td>05</td>
<td>花果山</td>
<td>5/0</td>
<td>17</td>
<td>玉皇大帝</td>
<td>7/0</td>
<td>29</td>
<td>吃月亮</td>
<td>1/4</td>
</tr>
<tr>
<td>06</td>
<td>杨二郎</td>
<td>6/0</td>
<td>18</td>
<td>毕马温</td>
<td>1/5</td>
<td>30</td>
<td>蟠桃会</td>
<td>5/1</td>
</tr>
<tr>
<td>07</td>
<td>大闹天宫</td>
<td>8/0</td>
<td>19</td>
<td>骆驼</td>
<td>4/0</td>
<td>31</td>
<td>仙女</td>
<td>5/1</td>
</tr>
<tr>
<td>08</td>
<td>笑天犬</td>
<td>2/4</td>
<td>20</td>
<td>李靖</td>
<td>2/4</td>
<td>32</td>
<td>定身法</td>
<td>6/1</td>
</tr>
<tr>
<td>09</td>
<td>天狗</td>
<td>3/5</td>
<td>21</td>
<td>托塔天王</td>
<td>3/0</td>
<td>33</td>
<td>巨灵神</td>
<td>4/1</td>
</tr>
<tr>
<td>10</td>
<td>吃月亮</td>
<td>1/4</td>
<td>22</td>
<td>哪吒</td>
<td>7/0</td>
<td>34</td>
<td>火眼金睛</td>
<td>6/0</td>
</tr>
<tr>
<td>11</td>
<td>四大天王</td>
<td>5/0</td>
<td>23</td>
<td>三太子</td>
<td>1/3</td>
<td>35</td>
<td>太上老君</td>
<td>6/0</td>
</tr>
<tr>
<td>12</td>
<td>天兵天将</td>
<td>5/0</td>
<td>24</td>
<td>风火轮</td>
<td>7/1</td>
<td>36</td>
<td>虾兵蟹将</td>
<td>5/1</td>
</tr>
</tbody>
</table>

Table 2. The simplified tag set of the cartoon.

<table>
<thead>
<tr>
<th>SN</th>
<th>Tag</th>
<th>Valid Status</th>
<th>SN</th>
<th>Tag</th>
<th>Valid Status</th>
<th>SN</th>
<th>Tag</th>
<th>Valid Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>孙悟空</td>
<td>1 (8÷0=16)</td>
<td>09</td>
<td>龙王</td>
<td>1 (5÷0=10)</td>
<td>17</td>
<td>蟠桃会</td>
<td>1 (5÷1= 5)</td>
</tr>
<tr>
<td>02</td>
<td>太白金星</td>
<td>1 (8÷0=16)</td>
<td>10</td>
<td>东海龙宫</td>
<td>1 (4÷0= 8)</td>
<td>18</td>
<td>仙女</td>
<td>1 (5÷1= 5)</td>
</tr>
<tr>
<td>03</td>
<td>花果山</td>
<td>1 (5÷0=10)</td>
<td>11</td>
<td>玉皇大帝</td>
<td>1 (7÷0=14)</td>
<td>19</td>
<td>定身法</td>
<td>1 (6÷1= 6)</td>
</tr>
<tr>
<td>04</td>
<td>杨二郎</td>
<td>1 (6÷0=12)</td>
<td>12</td>
<td>毕马温</td>
<td>1 (4÷0= 8)</td>
<td>20</td>
<td>巨灵神</td>
<td>1 (4÷1= 4)</td>
</tr>
<tr>
<td>05</td>
<td>大闹天宫</td>
<td>1 (8÷0=16)</td>
<td>13</td>
<td>托塔天王</td>
<td>1 (3÷0= 6)</td>
<td>21</td>
<td>火眼金睛</td>
<td>1 (6÷0=12)</td>
</tr>
<tr>
<td>06</td>
<td>四大天王</td>
<td>1 (5÷0=10)</td>
<td>14</td>
<td>哪吒</td>
<td>1 (7÷0=14)</td>
<td>22</td>
<td>太上老君</td>
<td>1 (6÷0=12)</td>
</tr>
<tr>
<td>07</td>
<td>天兵天将</td>
<td>1 (5÷0=10)</td>
<td>15</td>
<td>风火轮</td>
<td>1 (7÷1= 7)</td>
<td>23</td>
<td>虾兵蟹将</td>
<td>1 (5÷1= 5)</td>
</tr>
<tr>
<td>08</td>
<td>哮天犬</td>
<td>1 (5÷0=10)</td>
<td>16</td>
<td>水帘洞</td>
<td>1 (6÷1= 6)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. The final tag set of the cartoon.

<table>
<thead>
<tr>
<th>RN</th>
<th>Tag</th>
<th>Approving Ratio</th>
<th>RN</th>
<th>Tag</th>
<th>Approving Ratio</th>
<th>RN</th>
<th>Tag</th>
<th>Approving Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>大闹天宫</td>
<td>16</td>
<td>09</td>
<td>花果山</td>
<td>10</td>
<td>17</td>
<td>定身法</td>
<td>6</td>
</tr>
<tr>
<td>02</td>
<td>孙悟空</td>
<td>16</td>
<td>10</td>
<td>龙王</td>
<td>10</td>
<td>18</td>
<td>水帘洞</td>
<td>6</td>
</tr>
<tr>
<td>03</td>
<td>太白金星</td>
<td>16</td>
<td>11</td>
<td>四大天王</td>
<td>10</td>
<td>19</td>
<td>托塔天王</td>
<td>6</td>
</tr>
<tr>
<td>04</td>
<td>哪吒</td>
<td>14</td>
<td>12</td>
<td>天兵天将</td>
<td>10</td>
<td>20</td>
<td>蟠桃会</td>
<td>5</td>
</tr>
<tr>
<td>05</td>
<td>玉皇大帝</td>
<td>14</td>
<td>13</td>
<td>哮天犬</td>
<td>10</td>
<td>21</td>
<td>虾兵蟹将</td>
<td>5</td>
</tr>
<tr>
<td>06</td>
<td>火眼金睛</td>
<td>12</td>
<td>14</td>
<td>毕马温</td>
<td>8</td>
<td>22</td>
<td>仙女</td>
<td>5</td>
</tr>
<tr>
<td>07</td>
<td>太上老君</td>
<td>12</td>
<td>15</td>
<td>东海龙宫</td>
<td>8</td>
<td>23</td>
<td>巨灵神</td>
<td>4</td>
</tr>
<tr>
<td>08</td>
<td>杨二郎</td>
<td>12</td>
<td>16</td>
<td>风火轮</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion

Summary extract of heterogeneous multimedia documents is an important content and requirement of information management at this era. Our approach is reasonable and effective. There are some characteristics of the proposed approach: simplicity, economy, self-learning and dynamic evolution. In the future, we will further improve the efficiency of the approach.

For improving the efficiency, we conceive two strategies. The first one is building a similarity network. In a similarity network, the nodes (vertices) are all the tags whose valid status is 1, and if the similarity between two tags is bigger...
than a threshold determined by a group of experts or some kind of mechanism, then links an edge between the two tags. We can firstly conduct community detection, and then mine the star node in each community in such a network [32]. The star nodes and the isolated nodes in the network form a simplifying tag set. The second one is only reserving the tags which are in the top 20% tag ranking of the final tag set according to the Pareto principle.

Acknowledgment

This research was financially supported by the Heilongjiang Education Department Science and Technology Research Program (No. 12531529).

References


