Finding Semantically Related Images in the WWW

Heng Tao Shen Beng Chin Ooi Kian-Lee Tan
Department of Computer Science
National University of Singapore

1. INTRODUCTION

In this demonstration, we present a system designed to find semantically relevant images that are embedded in HTML documents in the WWW. The system has been implemented in Java on a Sun Sparc machine, and our experimental study showed the effectiveness of the system [1].

2. IMAGE REPRESENTATION MODEL

The main observation that we made is that an embedded image's semantics are typically captured by its surrounding text in the document. We have identified four parts of the textual content that are well related to the embedded image. These are the *image title, image ALT (alternate text), image caption* and *page title.*

To represent the image semantics more adequately, we propose the Weight ChainNet model that is based on the concept of *lexical chain*. A lexical chain (LC) is a sequence of semantically related words in a text. Here, we define it as one sentence that carries certain semantics by its words. As an image title is just a single word, we say it's a trivial lexical chain - *Title Lexical Chain (TLC).* The text obtained from the ALT tag is referred to as the *Alt Lexical Chain (ALC).* The page title is represented as a LC too - *Page Lexical Chain (PLC).* Finally, since a caption comprises multiple sentences, we represent it as three types of lexical chains. Type one is called *sentence lexical chain (SLC)*, which represents one single sentence in an image caption. Type two is called *reconstructed sentence lexical chain (RSLC)*, and it represents one new sentence reconstructed from related sentences. Two sentences are *related* if both share one or more words. One common word in two SLCs splits each SLC into two. Based on the first common word, the second SLC's second half is connected to the first SLC's first half to form a RSLC. The last type is called *caption lexical chain (CLC)*, which represents the whole image caption. A CLC is formed by connecting SLC one after another.

The ChainNet model is built with these 6 types of lexical chains. To capture the relative importance of the various types of LCs, we assign weights to the various LCs such that LCs that are deem to be more representative of the image content are assigned larger weight values.

For a user query, it's usually a free sentence that describes the image content. Naturally, we represent it as a *Query Lexical Chain - QLC*.

The similarity measure used to determine the degree of similarity between QLC and the various types of LCs of a database image essentially matches each type of LCs against the QLC using the list-space model. However, there are several novel features. First, as mentioned, LCs that are more critical are given higher weights. Second, we capture the closeness of two LCs from the view of match order. For example, one LC is " *US president Clinton and wife visited China in 1997*", and the other one is: " *China president Jiang Zemin welcomed Clinton and wife in Tian'an square*". For these two LCs, there are four matching words. For the first LC, the matched words are in order of "*president Clinton wife China",* and in the other, they are "*china president Clinton wife"*. We treat each one as a child LC of its original LC. Therefore, the orders of matched words in the two original LCs are not the same. Obviously, the closer the matched order of two children LCs are, the closer the semantics of the original two LCs are. Third, to ensure that an LC and the QLC are *semantically related*, we require the pair shares a certain minimum number of *distinct* matched words.

3. RELEVANCE FEEDBACK

Because of the large image collection and the impreciseness of a query, it is important to provide mechanisms to help users to specify their queries more accurately. For this purpose, we have also developed two feedback techniques:

Semantic Accumulation. In *semantic accumulation*, the user picks one *most relevant* image (from the user's subjective judgement) from the results of previous retrieval as the feedback image. The method accumulates all the previous feedback images' semantics to construct a new query for the next retrieval. The resultant query is represented as a kind of ChainNet called Weight F/Q ChainNet (Feedback/Query ChainNet) since it is constructed by the query and the feedback image's ChainNet.

Semantic Integration and Differentiation. In this method, users can select several relevant and irrelevant images simultaneously. By relevant, we mean images that are semantically related to the query as judged by the user and hence should be retrieved. On the other hand, irrelevant images are those that the user considers to be unrelated and should not have been retrieved. The system *integrates* the related semantics obtained from the relevant feedback images to construct a new query for the next try. After that, the system combines the semantics from irrelevant images to *differentiate* the irrelevant images from the returned results.

4. SYSTEM DESIGN

Figure 1 shows the architecture of the system. The system consists of five basic components: a *Web Crawler*, a *QLC Generator*, an *Image ChainNet Generator*, an *Image Search*

Engine, and a *F/Q ChainNet Generator*. The database consists of three parts: WordNet, Query Profile, and Image Database.

The Web Crawler that operates in the background automatically searches the WWW for documents with embedded images. The crawler also extracts the image title, image ALT, page URL, page title and image caption from the HTML documents as the images' semantic content. It also loads the meaningful images (or their URLs) with their representation into the database. For purpose of testing the model, we "centralized" the image collection (instead of simply extracting the image at runtime from the various Web sites/pages in the form of a search engine). The QLC Generator transforms a user query (free text description) to a *query lexical chain* (QLC). The Image ChainNet Generator constructs image content semantic representation: image weight ChainNet. It retrieves image content from database and creates ChainNet object. The Image Search Engine compares the QLC against the ChainNet of the images, and returns all semantically related images. The images are displayed in order of decreasing degree of similarity. Finally, the F/Q ChainNet Generator is used to generate extended query from the user query and the feedback images. Given the selected images from users, the F/Q ChainNet Generator creates a new weight ChainNet by combining the selected image's ChainNet, and the QLC. Depending on the feedback mechanisms adopted, different new weight ChainNet may be obtained. Finally, the Image Search Engine will perform the next round of retrieval based on the output of F/Q ChainNet Generator.

5. STRUCTURE OF THE DEMO

In this demonstration, we shall illustrate the effectiveness of the proposed system in the following ways:

1. **Semantic extraction**. Given a document with an embedded image, we shall extract its ChainNet. This allows us to have a feel of the exactness of the semantic representation.
2. **Image retrieval**. Given a query (i.e., a set of keywords), we find images that are semantically related to the query. To demonstrate the effectiveness of the proposed system, we shall examine two different systems: traditional approach that is based purely on keyword matching and proposed approach that is based on the Weight ChainNet model.
3. **Feedback Mechanisms**. Given a query, we shall retrieve its semantically relevant images. In addition, the two feedback mechanisms will be employed to facilitate more precise retrieval. This demonstration will also show the relative differences (pros and cons) of the two proposed methods. In the process, we will also show the revised query as a result of the feedback.
4. **Effects of Weights**. The effectiveness of the system is dependent on values assigned to the weights. For the purpose of the demo, the system has been designed so that the tuning parameters can be set "externally". This demonstration will allow us to "tune" the weights, and observe their effects. In fact, we shall see that the default picked by us (obtained after extensive experimental study) is indeed reasonably optimal. Setting the weights externally also has the advantage of allowing different emphasis on different lexical chains.

References

[1] H.T. Shen, B.C. Ooi, K.L. Tan, *Giving Meanings to WWW Images.* MM'00 (to appear).

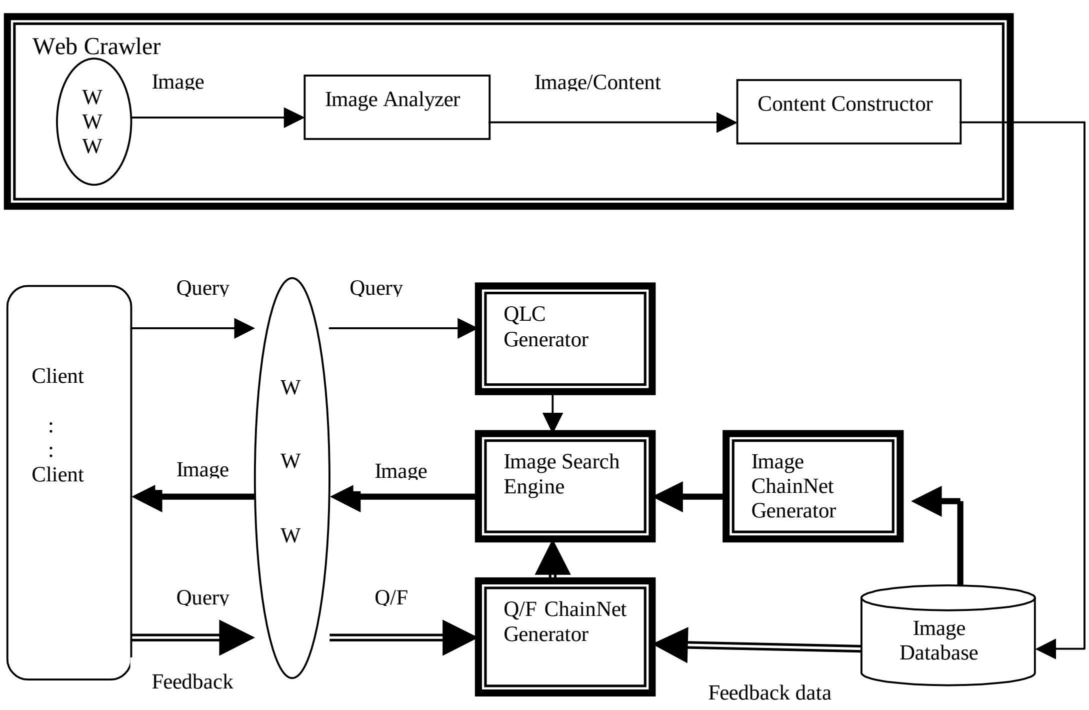

Figure1: Overall system structure in client-server form.