# Designing A Question-Answering System for Comic Contents

Yukihiro Moriyama
Kansai University
Takatsuki, Osaka, Japan
k700331@kansai-u.ac.jp

Byeongseon Park
Kansai University
Takatsuki, Osaka, Japan
k281401@kansai-u.ac.jp

Shinnosuke Iwaoki
Kansai University
Takatsuki, Osaka, Japan
k372123@kansai-u.ac.jp

Mitsunori Matsushita
Kansai University
Takatsuki, Osaka, Japan
m_mat@kansai-u.ac.jp

## ABSTRACT

The objective of our research is to create a question answering system for comics. Because a comic has multimodal contents, we have to answer questions about text as well as illustrations. This is different from the conventional question answering system. To solve this problem, in this study, we organized the information to be obtained from comic illustrations and examined the framework of question answering for this content. Then, we built a prototype system and examined the question answering system for comic contents.

## CCS Concepts

•**Information systems** → **Question answering;**

## Keywords

comic computing; question answering technology; query classification;

## 1. INTRODUCTION

With the proliferation of electronic devices such as tablets and smart phones, it has become popular for people to read comics on such devices. In line with that, the market for digital comics has grown rapidly in recent years. In Japan, for instance, the sales amount of digital comics in 2015 rose to 1,149 billion Japanese Yen, up 38.1% from the previous year. With such an increase in the availability of digital comics, the desired comic book content can be found more efficiently. Digital comics are regarded as having higher applicability than paper comics because digitalized content is easy to manipulate dynamically with computers and smart devices.

Currently, a user can retrieve a comic by searching publishers' web pages (e.g. Kodansha Comic Plus[1]) or a bookstore (e.g., Comic-Cmoa[2]). Search results can be based on bibliographic information (e.g., title and author) or genre information (e.g., fantasy, humor, and Sci-Fi). However, comic readers' interests regarding comic book contents have been changing along with the spread of the digital comics. For instance, users may want to examine the outline from the comic, or get character descriptions for comics (e.g. the main character is a feeble grappler) or the themes (e.g. battle, love story). Moreover, the user may want to know more detailed information (e.g. "What is the volume in One Piece that Chopper becomes a friend with Ruffy?").

In the face of such requirements, current search function offered by publishers and bookstores falls short, and users often consult QA sites including knowledge sharing community services (e.g., Yahoo Chiebukuro[3] and Oshiete Goo[4]) or a FAQ list on the comic's fun site (e.g., One Piece Fun Site[5]).

The QA sites are useful, however, there are cases when the answer is not accurate, or it may take time to obtain the answer, because the answers are made by hand in response to the user's question. In addition, if a user wants to know the number of occurrences of the character or short summary of the story, it is difficult to satisfy the user's need because the answer cannot usually be obtained on a QA site.

Based on these limitations, we intend to develop a question answering system for comic contents (Comic QA) in which answers are generated from the comic contents automatically. The comic QA will enable users to obtain information directly based on the comic contents including episodes and/or in the story, the characters, and other characteristics of the comic. To meet this goal, this study investigates types of information to be extracted from the comic panels, and proposes a basic framework for constructing the Comic QA. In addition, we develop a prototype of Comic QA system with the information and review feasibility of the system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

*MANPU '16, December 04 2016, Cancun, Mexico*

© 2016 ACM. ISBN 978-1-4503-4784-6/16/12…$15.00

DOI: http://dx.doi.org/10.1145/3011549.3011554

[1]http://kc.kodansha.co.jp/search (Confirmed 1, Sep. 2016).
[2]http://www.cmoa.jp/ (Confirmed 1, Sep. 2016).
[3]http://chiebukuro.yahoo.co.jp/ (Confirmed 1, Sep. 2016).
[4]http://oshiete.goo.ne.jp/ (Confirmed 1, Sep. 2016).
[5]http://onep.jp/modules/onepwiki/ (Confirmed 1, Sep. 2016).

## 2. RELATED WORK

### 2.1 Realization of Various Question Answering Systems

Jacquemart et al. studied the feasibility of a question answering system for health care. They collected a corpus of student questions in oral surgery. Moreover, they examined two conditions: (1) how to select the right keywords in a question to identify relevant material on the Web, (2) whether the contents of the question fit the conceptual model of an existing QA prototype[4]. Cruchet et al. also studied the approach to recognize question type in a question answering system for health. Their original hypothesis for the system is that a question can be defined by two criteria: type of answer expected and medical type. They collected questions from the Internet. The questions were given to medical experts in order to obtain a classification based on the certain criteria. Based on their results they proposed and evaluated a classification algorithm based on the result[2]. Kwok et al. proposed and evaluated a fully automated question answering system that is available on the web in order to answer to factual questions such as "who was the first American in space?" or "what is the second tallest mountain in the world?"[5]. In addition, Garcia et al. proposed a question answering systems based on the semantics. The system takes queries expressed in natural language and an ontology as input and returns answers drawn from the available semantic markup[3]. Hanxing et al. also proposed ontological question answering system based on FAQ in order to realize a system based the semantics[6].

### 2.2 Image Retrieval

There is text-based image retrieval (TBIR) and content-based image retrieval (CBIR) roughly in image retrieval area. First, the TBIR method searches the image matched input query and text data linked to images. For instance, Zhang et al. proposed an image retrieval system that returned the image that matched the search query and images surrounding text obtained from various photo forum web sites [10]. Yee et al. proposed an image retrieval system based on a facility that gave meta information to each image in a large number of image groups [9]. The CBIR method searches the image for a feature that is similar to a feature in the query image. For instance, Zhu et al. proposed an image retrieval method based on color and texture as the feature [11]. In addition, Belongie et al. proposed a retrieval method utilizing shape context information as the feature [1]. On problem associated with Content-based Image Retrieval is the problem so-called "semantic gap". A "semantic gap" can occur when the retrieved image is not visibly similar to the image used in the query. To solve the problem, there are studies that combine Text-based and Content-based Image Retrieval. For instance, Wang et al. proposed a method of annotating images by mining images and the surrounding text of the search results [8].

## 3. QUESTION ANWERING FOR COMICS

### 3.1 Organizing Comic Contents

Comic can be regarded as a multimodal content that constitutes texts (e.g., script and onomatopoeia) and illustrations (e.g., character, tool and background). In a comic, a page is divided into panels, with drawn texts and illustrations in each panel. Moreover, the text often is placed as the illustration in the content to vary from the text mainly of medium such as newspaper articles, which is various positions and the shapes. Thus, comic content composed of text and illustrations that are complementary and cooperative.

There is a variety of information users may seek about a comic, such as, the number of volumes of the comic (e.g., "in what volume does the content I want to see occur?"), the summary of the story (e.g., "I want to know what happens at the end of the story."), and the description of a character in comic (e.g., "What skills does the main character have?"). It is necessary to understand the search technique in order to know whether to focus on the bibliographic information or the content information of the comics so that can be answered appropriately. However, it is difficult to apply conventional document retrieval since comic content is a mixture of text and illustrations. Further, it is necessary to organize a comic ' s content prior to performing a search. Morozumi et al. organized comics for the purpose of flexible information access on the comics by refering to the entity model of FRBR [7]. This study is useful for accessing bibliographic information such as volumes and pages. For the realization of this study, it is also necessary to organize comics based on both types of content information.

### 3.2 Information Requirements about Comics

The information requirements of comics include requirements about bibliographic content (e.g., I want to seek Sci-Fi comic) and/or the content in the comic (e.g., I want to see the scene in the comic). Bibliographic information of comics can be acquired by referring to the web sites, such as Wikipedia and the HP of published magazines. However, it is difficult to automatically collect content information of comics from web sites. For instance, if a user wants to know the volume of Dragon Ball where the scene where Kuririn uses a Kamehameha is depicted, the user needs to read the comics from the beginning and/or seek the required information from web sites. As comic content is made up of text and illustrations, it is difficult to accumulate text information for information retrieval. Users attempt to resolve this issue by using QA sites on the web when he/she wants to know about the content of the comic. However, the user cannot gather accurate information immediately from the question site because answers are written by unspecified users. There are cases of giving a wrong answer to a question or taking a long time to reply. Comic QA enables users to see only the information that answers their questions input using natural language. Consequently, Comic QA can save users a lot of trouble of reading the comics from the beginning and/or seeking the required information in the web sites, present the appropriate information immediately to them to solve the problem in QA site during use. To achieve this, it is necessary to consider the following: (1) Appropriate information presentation method for the user, (2) Information retrieval techniques for realizing the presentation method.

## 4. SYSTEM DESIGN

In the section, we discuss the information required for Comic QA by referring to the process from inputting a question to presenting an answer in existing question answering systems. Moreover, we proposes the system architecture for Comic QA based on the matters.

## 4.1 Process of Existing QA System

Existing question answering systems searches answer documents by using a user's question such as a query from the web, extracts named entities from the documents, present the appropriate information from the named entities. In this study, we organized the following three basic processes of Comic QA referring to the process from inputting questing to presenting answer in existing question answering systems.

**(1) Question Analysis**
The process analyzes interrogative (e.g., "When", "Why") and around the words in a question inputted by a user, classifies the question sentence in order to decide presenting information as answering to the question. Moreover, this process extracts independent words (e.g., noun, adjective) as queries in order to retrieve and extract information retrieval.

**(2) Information Retrieval and Extraction**
The process searches a document likely to contain the answer using the queries and extracts the answer (named entity) from the document.

**(3) Answer Presentation**
The process presents answers from the named entities in order of rank. The information is not necessarily correct in case documents include a variety of topics. The process increases the score of the information close semantic distance between a query and them, presents the top information in the ranking as the system output.

## 4.2 Process of Comic QA System

As mentioned in Section 3.1, comic has multimodal contents that comprise texts and illustrations, hardly describe the text information about description of story, page, panel. It is necessary to process a comic by focusing on the specific structure of comic contents. In this paper, we discuss the following three points in order to develop Comic QA.

**(1) Interpreting Question Sentence for Comic**
As a user inputs a query by using natural language in Comic QA, it is possible to apply the type classification method for framing the sentences in a question in existing question answering systems. However, it is necessary to discuss various types of sentences in a question for comic and clue words for retrieving information in Comic QA. Comics have specific expressions, words, and names. Comic QA needs to remember them in order to use them.

**(2) Information Retrieval and Extraction for Comic**

There is a significant difference regarding the structure of the content between existing question answering systems for text and Comic QA. Therefore, Comic QA requires information retrieval and extraction method that focus on the specific structure of comic contents. For example, a comics rarely explain the elements depicted as image information (e.g., secret tool of "Doraemon", devil fruit of "One Piece") in comic panels by using text, as users see and understand the illustration. It is difficult to apply existing information retrieval and extraction methods when a user wants to know elements not represented in the text. Consequently, Comic QA needs to respond to users' needs by recognizing the complex structure of comics. Information retrieval and extraction in Comic QA needs to organize comic contents, create the data set to acquire flexibly the information from clue of question sentences.

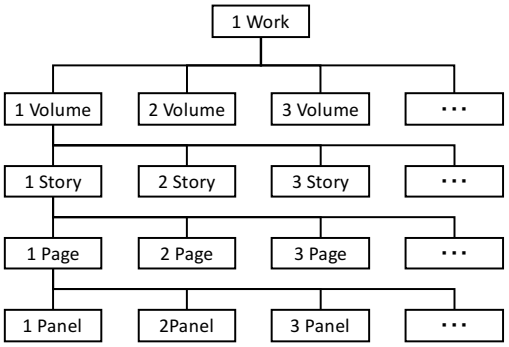


**Figure 1: Physical Hierarchical Structure of Comic**

**(3) Selection Answer in Comic QA**
The process followed in existing question answering systems presents named entity for users. Comic QA might present comic images better than the named entity as answer. Therefore, it is necessary to discuss answering presentation method using image data for users.

# 5. PROTOTYPE SYSTEM

## 5.1 Data Set

Our developed system uses the secret tools listed in volumes 1 to 7 of "Doraemon". The system is able to present three types of information to users: (1) appearance position type (e.g., volume number and story number of a depicted tool), (2) description type (e.g., "Take-copter is the secret tool to fly"), (3) scene type (e.g., use situation image of the secret tool). The three data sets were organized manually in order to find the answer by using keywords from the question.

First, a data set was organized into volume numbers, story numbers, page numbers, panels by referring to the physical hierarchical structure of comic (see Figure 1) [7]. Panels are the lowermost layer in a structure. The data set includes elements (e.g., secret tools) that appear in panels, as shown as in Table 1. Thus, the system is able to find appearance position of elements in each panel (DB1 in Figure 3). Second, we prepared the panel image by using a secret tool in comic by hand (DB2 in Figure 3), and secret tool's text information by hand by referring to dictionary sites (e.g., Wikipedia, Wikia) in order to find the description about the secret tool (DB3 in Figure 3). Thus, the system is able to find information that matches a secret tool's name.

This study develops a prototype for the question answering system for comic contents using the data sets, as described above. Figure 2 shows the system interface. When users input a question sentence in the input screen (Figure 2-a) and click the send button (Figure 2-b), the corresponding answer is displayed on the screen (Figure 2-c).

Table 1: A Part of Appearance Position's Data

| Volumes | Stories | Pages | Panels | Secret Tools |
|---|---|---|---|---|
| 2 | 9 | 98 | 1 | |
| 2 | 9 | 98 | 2 | Time-Furoshiki |
| 2 | 9 | 98 | 3 | Time-Furoshiki |
| 2 | 9 | 98 | 4 | Time-Furoshiki |
| 2 | 9 | 98 | 5 | |
| 2 | 9 | 98 | 6 | Time-Furoshiki |
| 2 | 9 | 98 | 7 | Time-Furoshiki |
| 2 | 10 | 99 | 1 | |

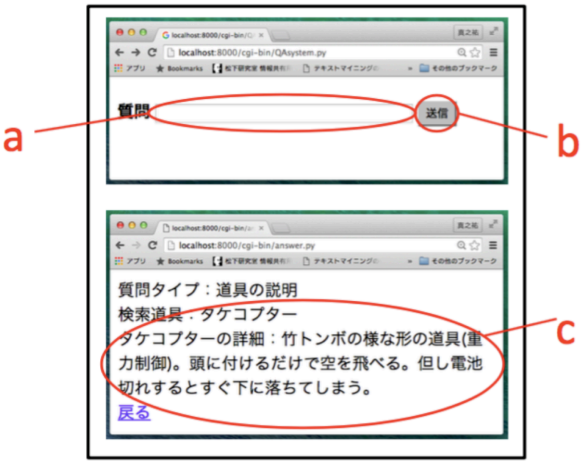


**Figure 2: Interface of Our Developed Comic QA. (Translation of c: "*Query-type:*Description of tool, *Retrieved tool:*Hopter, *Tool details:*Hopter is a tool that looks like a Japanese traditional toy, Bamboo-copter. This tool can control gravity. Users can fly by simply attaching it on the head. It does not work immediately after energy becomes empty.")**

## 5.2 Process

Figure 3 shows the overview of the question answering system for comic.

**(1) Question Sentence Analysis**

Question sentence analysis uses the morphological analyzer for Japanese text, Mecab. However, there are cases where the words in comic are not recognized as one morpheme. For example, Mecab might recognize a secret tool "Spy-Set" as "spy" and "set" without recognizing the tool as one word. We registered words to be recognized as one morpheme in the dictionary a priori in order to prevent this problem.

**(2) Question Type Classification**

We collected a corpus based on each question type as learning data for measuring the similarity between documents in order to classify the three question types automatically by machine learning. The system use cosine similarity to compare the documents to each other in order to classify the question sentences. Figure 4 shows the configuration of the type classification. This proposed system compares inputted the question sentence and each corpus. The most similar corpus

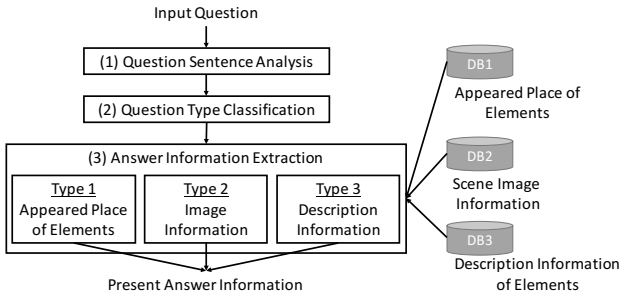


**Figure 3: Process of Question Answering System for Comic**

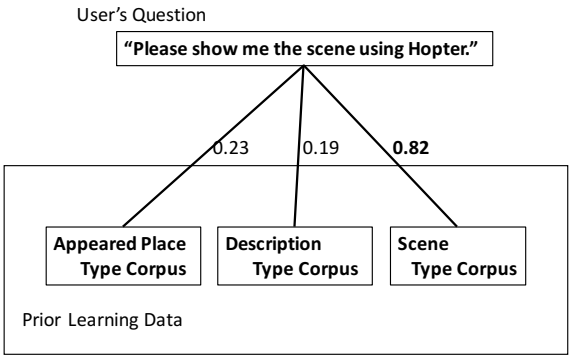


**Figure 4: Example of Question Type Classification**

is a type of the question sentence. The type of question sentence determines the question about scene type with the highest degree of similarity between the question sentence and the corpus about the scene type, as shown in Figure 4.

**(3) Answer Information Extraction**

The system extracts the secret tool as a search word from the question sentence by analyzing the morpheme using MeCab. The process uses different information extraction methods for each question type. The appearance position type extracts the position (e.g., volume number) of the secret tool that corresponds to the keyword. The description type extracts the text document about the secret tool. The scene type extracts the image data about the use situation of the secret tool.

The system uses a different presentation method for each question type in order to appropriately present the extracted information to the users. The presentation screen displays the classified question type and the secret tool, which is a search word, as common information.

## 5.3 Execution Result

The system uses different answering presentation methods for each question type. The answering screen displays the classified question type and the search word. Appearance position type's answer presents text information where the secret tool appeared (Figure 5). The description type's an-

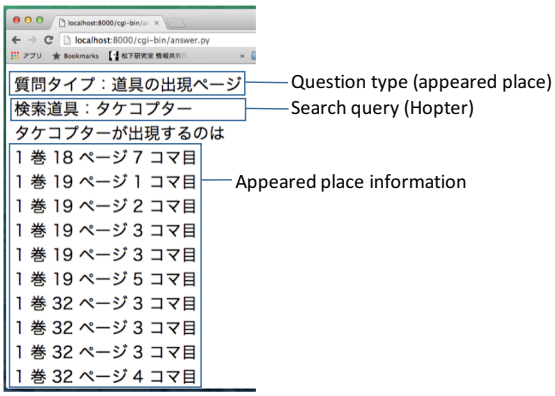


**Figure 5: Example of Answer to Appearance Position Type Question**

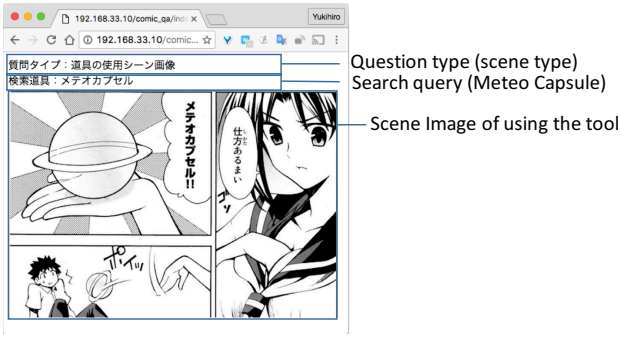


**Figure 6: Example of Answer to Scene Type Question ©Takuji from Manga109**

swer presents text information and image data of comic page (Figure 2). Scene type's anwer presents image information (e.g., use situation of the tool) (Figure 6).

## 6. DISCUSSION

In this study, comic contents were categorized into volume numbers, story numbers, page numbers, and panels in order to find the appearance position of specific elements in comic. The content used in this study was clearly classified in to panels and was part of a simple story. The physical hierarchical structure of comic is applied in this case. However, other comics might use the panels as opposed to drawing the contents. For example, a comic inserts the white panel in order to clarify the character's orientation (Figure 8), and superimposes multiple white panels in order to express the character's feeling and lingering or draws the character on multiple panels in the page (Figure 7). Therefore, it is necessary to discuss these panel features and elements in the comic content. The implemented system limits the corresponding possible question sentence. It is difficult to correspond to various user information needs. In this paper, the required data is organized to answer to each question sentence by using the answers intended for about 200 of the question sentences from Yahoo! Answers and Oshiete Goo in order to clarify the data to enable association with the structured comic.

**(1) Multiple Panels or Pages**

The data about appearance position information is required to include the information that indicates a particular scene or story continued on multiple pages. If there is a description such as extra edition in the comic content, this information is often inputted as a clue in the question sentence.

If the information is not described and the scene is represented by illustrations, the question sentence comprises only ambiguous clues (e.g., cool scene, the success story of the main character). Therefore, it is necessary to discuss the compatible data and structure to these clues.

**(2) Relationships Between Characters**

Some of the collected question sentences included those about the relationship between the characters. For example, the answer to the question, "What kind of relationship do character A and character B share?" might create or utilize a correlation diagram of the characters. However, for the answer to the question, "Why character A and character B have developed such a relationship?", it is difficult to only display a correlation diagram of the characters. Therefore, it is necessary to utilize the data about causal relationships in order to answer to this question.

**(3) Each Content Definition**

This paper defined the object information of the important element "tool" in Doraemon in order to answer the question about description type. However, the questions for comic contents include skill definition, meaning of words, and character profiles in other comics. Therefore, it is necessary to define the information related to important elements of each content. Morover, it is necessary to define the information in order to answer to the question about things that does not be clearly drawn as object in the comic.

**(4) Comic Bibliographic Information**

Some of the collected question sentences included questions about the bibliography (e.g., "I want to look for comics by the author"). Therefore, it is necessary to define the data about bibliographic information such as author and publication date related to a comic's title in order to answer to the question. Moreover, there are the questions to request the comic title to the clue of comic contents. Therefore, it is necessary to associate the data about bibliographic information and the data about the comic contents.

**(5) Story Summary**

The study about comic summary considers the automatic summarization techniques to extract important panels from each story and each page. The comic summary might be described as the document information in Wikipedia and home page of the comic. Therefore, it is possible to present the information about comic summary by utilizing those techniques and source.

The question about comic summary included those about the story ending (e.g., "I want to know the end"). Therefore, it is necessary to discuss the summarization method that focuses on the story ending and organizes the data set for quick information retrieval.

**Figure 7: Scene Depicting the Character in Several Panels ©Yusuke Takeyama from Manga109**

**Figure 8: Insertion of Blank Panel Into the Panel in Order to Clarify the Character of the Orientation ©Yuzuru Shimazaki from Manga109**

## 7. CONCLUSION

This study attempted to develop a question answering system for comic contents. First, we discussed the information required for comic content by referring to the process used by the existing question answering systems. A comic has multimodal contents that comprise texts and illustrations. Therefore, it is difficult to apply the document retrieval techniques used in the existing question answering systems. Moreover, it is necessary to organize the data set focused on comic structure in order to search for comic contents. Second, the data set for the search was organized and a question answering system was developed for comic contents by utilizing the data to target the secret tools in Ladybug Comics "Doraemon" 1-7 Volume. The proposed system enabled users to request the question pertaining appearance position type, description type, and scene type related to the secret tool. Finally, the question sentences of other comic were analyzed and discussed the type of data set that the system should utilize, including the information retrieval technique, in order to realize a more generic question answering system.

## 8. ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 15H12103. Moreover, the data for the implemented system was utilized Shogakukan publication Ladybug Comics "Doraemon" (Fujiko · F · Fujio work).

## 9. REFERENCES

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.

[2] S. Cruchet, A. Gaudinat, and C. Boyer. Supervised approach to recognize question type in a QA system for health. *eHealth Beyond the Horizon – Get IT There*, 136:407–412, 2008.

[3] V. L. Garcia, E. Motta, and V. Uren. Aqualog: An ontology-driven question answering system to interface the semantic web. In *Proc. 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume: Demonstrations*, pages 269–272, 2006.

[4] P. Jacquemart and P. Zweigenbaum. Towards a medical question-answering system: a feasibility study. In *The New Navigators: from Professionals to Patients*, volume 95, pages 463–468. IOS Press, 2003.

[5] C. Kwok, O. Etzioni, and D. S. Weld. Scaling question answering to the web. *ACM Trans. Inf. Syst.*, 19(3):242–262, 2001.

[6] H. Liu, X. Lin, and C. Liu. Research and implementation of ontological qa system based on faq. *J. of Convergence Information Technology*, 5(3), 2010.

[7] A. Morozumi, S. Nomura, M. Nagamori, and S. Sugimoto. Metadata framework for manga: A multi-paradigm metadata description framework for degital comics. In *Proc. Int. Conf. on Dublin Core and Metadata Applications 2009*, pages 61–70, 2009.

[8] X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma. Annotating images by mining image search results. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(11):1919–1932, 2008.

[9] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pages 401–408, 2003.

[10] L. Zhang, L. Chen, F. Jing, K. Deng, and W.-Y. Ma. Enjoyphoto: A vertical image search engine for enjoying high-quality photos. In *Proc. 14th ACM International Conference on Multimedia*, pages 367–376, 2006.

[11] X. Zhu, J. Zhao, J. Yuan, and H. Xu. A fuzzy quantization approach to image retrieval based on color and texture. In *Proc. 3rd International Conference on Ubiquitous Information Management and Communication*, pages 141–149, 2009.