

Framework of Manga Application for Teaching Japanese Language

¹Ritsumeikan University , ²Kansai University , ³Simon Fraser University

Masahide Kuwano¹(is0305pk@ed.ritsumei.ac.jp), Ryosuke Yamanishi²(ryama@kansai-u.ac.jp),
Yoko Nishihara¹(nishihara@fc.ritsumei.ac.jp), Naoko Takei³(ntakei@sfu.ca)

Background & Purpose(1)

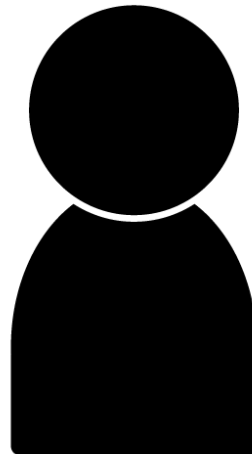
What kind of scene do you imagine from a sentence

“After dinner, the family shared an apple to eat.”?

In **Japan**, native Japanese naturally imagine a relatively big kind of apple “Fuji”.



Most of us Japanese imagine “A family circle.”



In **North America**, North Americans naturally imagine an apple small enough to fit in a hand.



North Americans may feel the opposite understanding like “The family is so poor that need to share an apple.”.

Background & Purpose(2)

The teacher read every page of the comics and find the speech including the intended words and phrases by his/herself.

→ The teacher spends a lot of time reading every page of the comics.

Also, the teacher may not find a speech that contains the intended words and phrases.

The teacher visually searches for an appropriate comic scene from one comic.



LoveHina © Ken Akamatsu

The teacher manually extracts the frames from the pages.



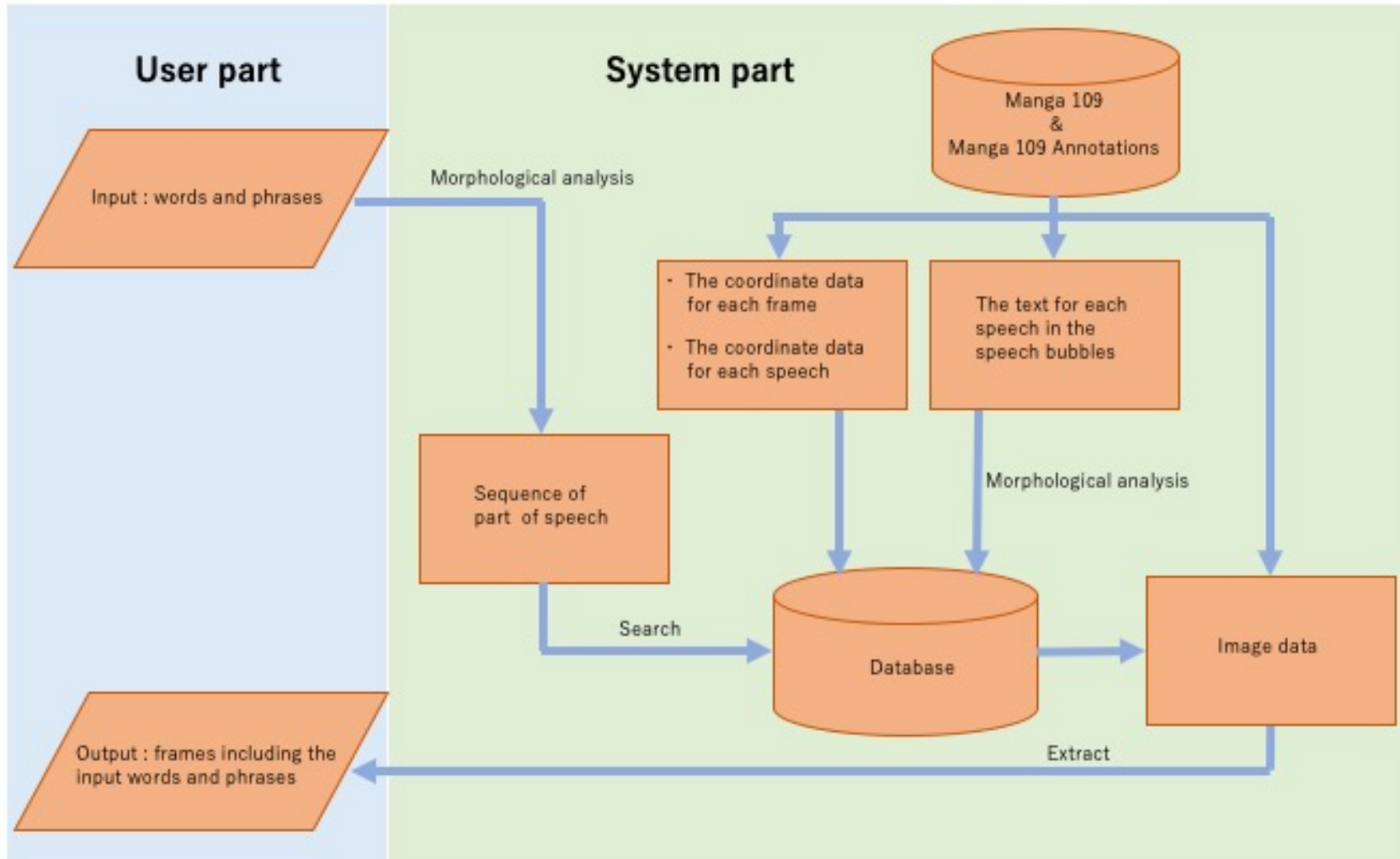
LoveHina © Ken Akamatsu



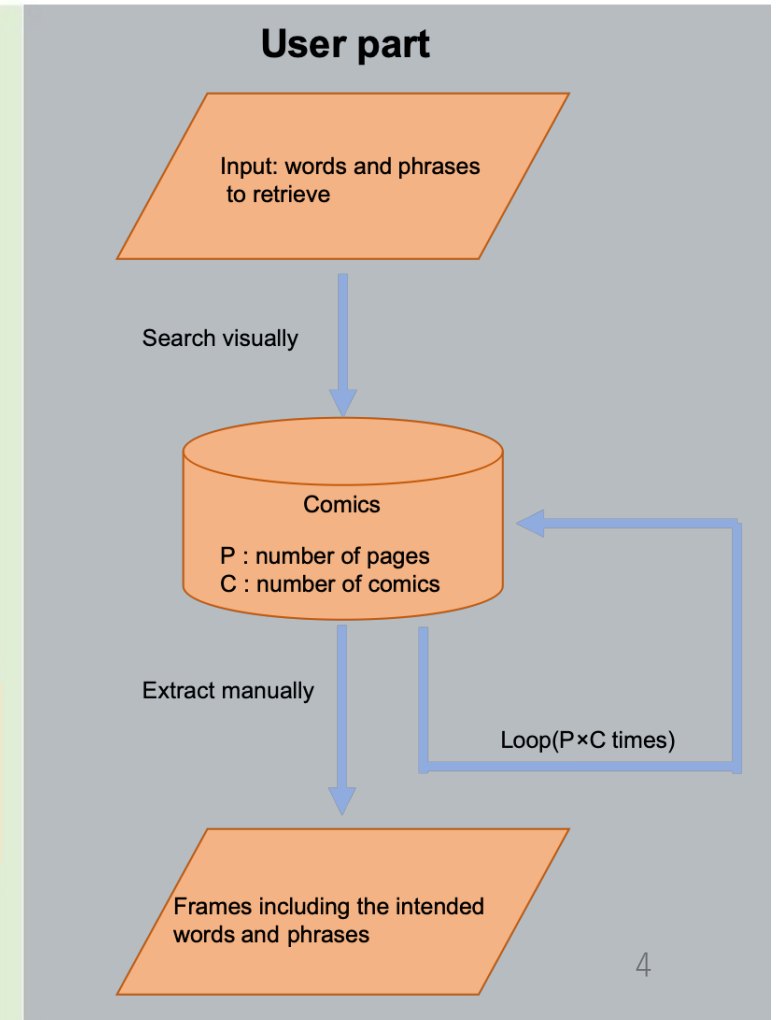
LoveHina © Ken Akamatsu

Comparison of traditional tasks and the proposed system architecture

The proposed system architecture



Traditional tasks



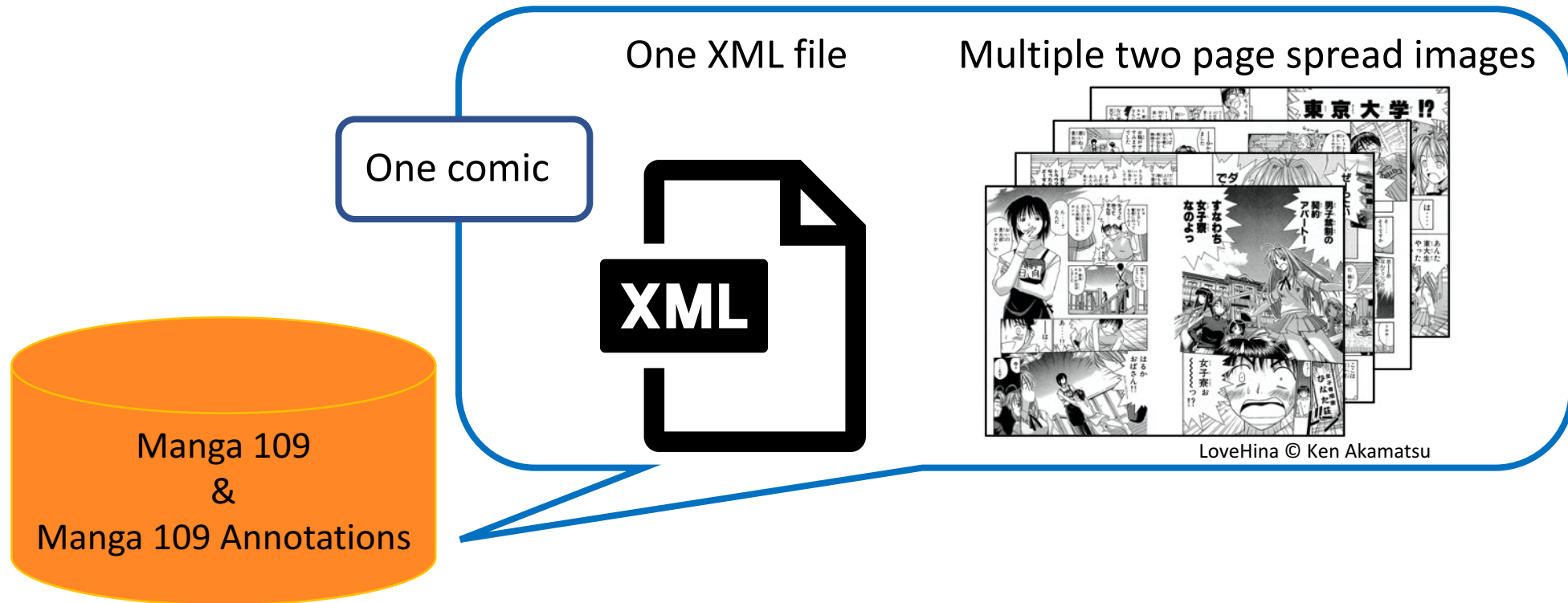
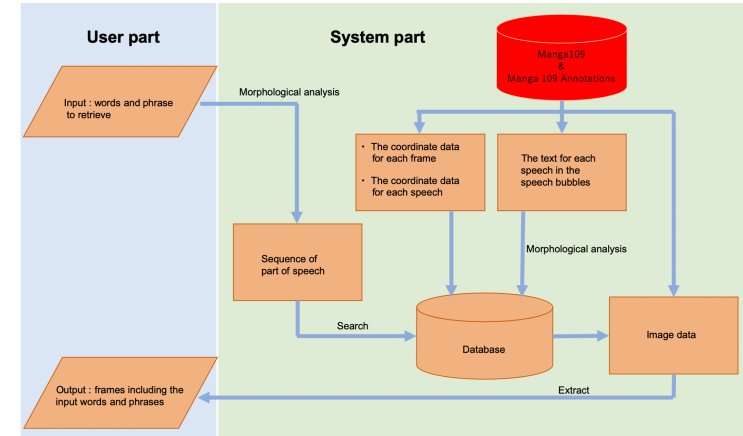
The process in the proposed system

1. The proposed system obtains annotations in the Manga 109 dataset.
2. The proposed system process the obtained annotations.
3. The proposed system process the input query from the user.
4. The proposed system outputs the frames including the query.

The proposed system obtains annotations in the Manga 109 dataset (1)

We used Manga 109 and Manga109 Annotations datasets.

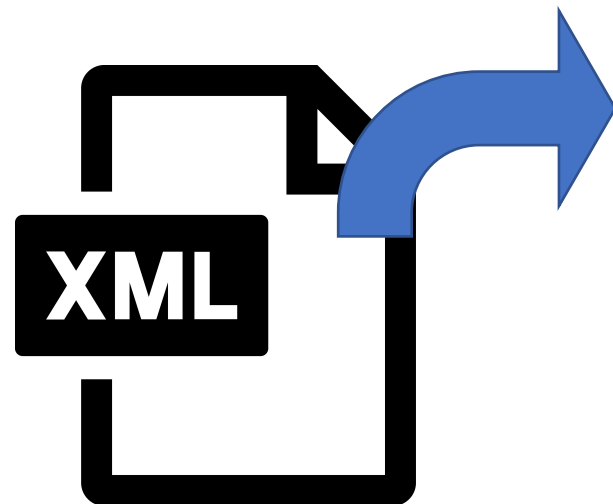
The dataset contains one XML file and multiple two pages spread images in one comic.



The proposed system obtains annotations in the Manga 109 dataset (2)

We used three types of annotations from one XML file in the Manga 109 dataset.

- The coordinate data of the speech bubble
- The text of the speech for the speech bubble
- the coordinate data of the frame

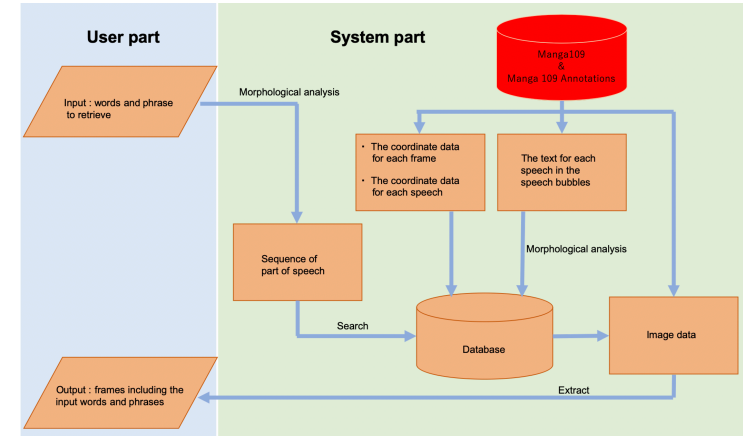


```
<body id="00036146" xmin="119" ymin="836" xmax="836" character="00036146"/>
<face id="00036147" xmin="35" ymin="890" xmax="890" character="00036147"/>
<face id="00036148" xmin="113" ymin="698" xmax="698" character="00036148"/>
<face id="00036149" xmin="68" ymin="776" xmax="776" character="00036149"/>
<face id="0003614a" xmin="1456" ymin="961" xmax="1493" ymax="998" character="00035f85"/>
<body id="0003614b" xmin="20" ymin="29" xmax="344" ymax="618" character="00036141"/>
<text id="0003614c" xmin="629" ymin="794" xmax="744" ymax="1003">はるかおばさん!! おひさしぶり--</text>
<text id="0003614d" xmin="913" ymin="60" xmax="1093" ymax="267">9なわら女子寮なのよっ</text>
<frame id="0003614e" xmin="908" ymin="780" xmax="1449" ymax="1167"/>
<body id="0003614f" xmin="314" ymin="837" xmax="837" character="00036141"/>
<face id="00036150" xmin="1054" ymin="837" xmax="837" character="00035faa"/>
<text id="00036151" xmin="649" ymin="837" xmax="837" character="00036151"/>おとなしく警察に行ってもらわよ</text>
<text id="00036152" xmin="931" ymin="837" xmax="837" character="00036152"/>女子寮お〜〜っ!?!</text>
<body id="00036153" xmin="156" ymin="837" xmax="339" ymax="1169" character="00035f83"/>
<text id="00036154" xmin="685" ymin="374" xmax="740" ymax="472">騒々しいなどうした</text>
<frame id="00036155" xmin="904" ymin="2" xmax="1651" ymax="832"/>
<text id="00036156" xmin="1457" ymin="64" xmax="1619" ymax="271">男子禁制の契約アパート!</text>
```

The coordinate data of the speech bubble

The text of the speech for the speech bubble

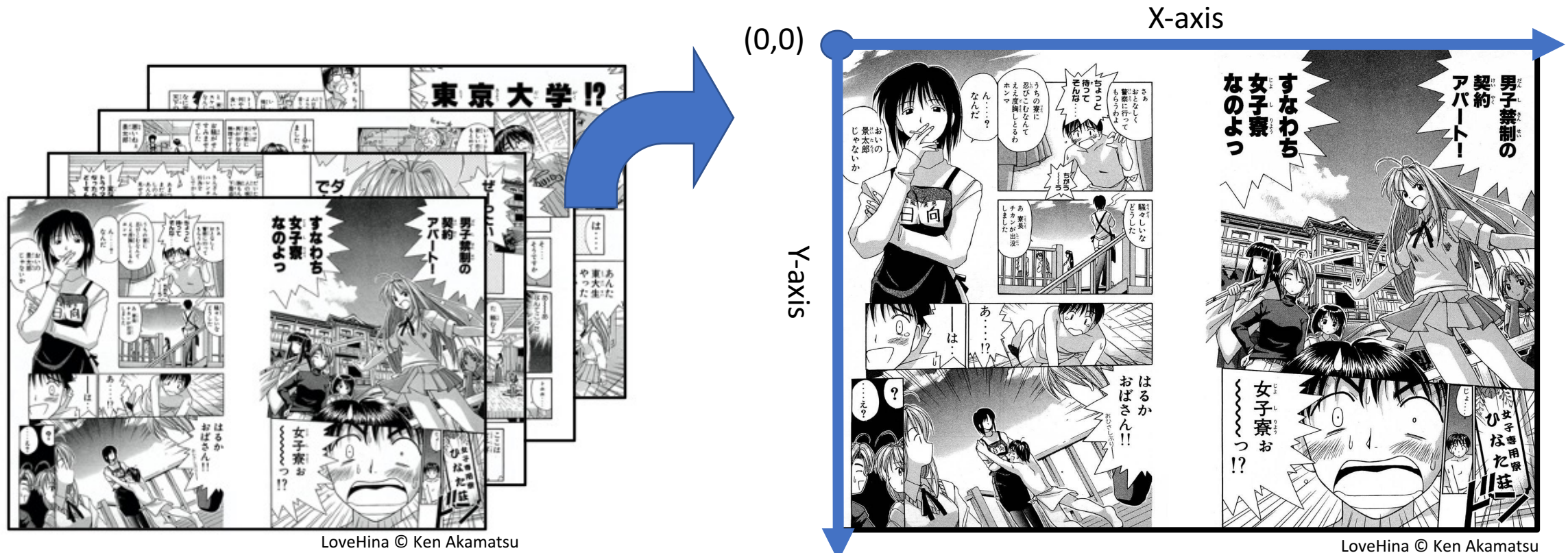
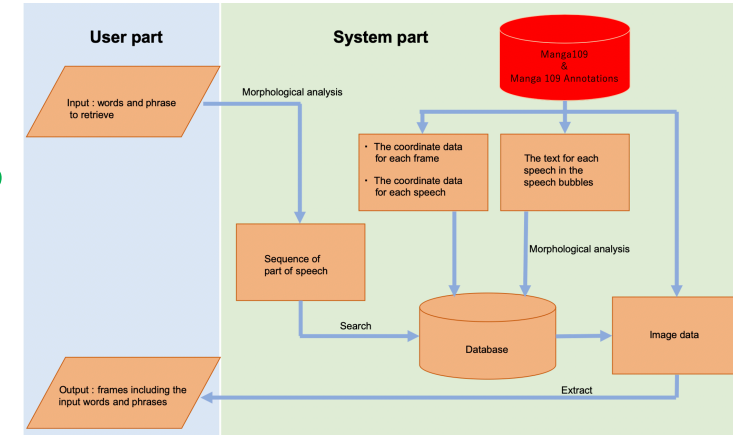
The coordinate data of the frame



The proposed system obtains annotations in the Manga 109 dataset (3)

We used multiple two pages spread images in the Manga 109 dataset.

- The horizontal axis of the image is the x-axis and the vertical axis is the y-axis.
- The upper-left coordinate of two pages spread image is (0,0).



The proposed system process the obtained annotations (1)

The proposed system processes the coordinate data of the speech bubble.

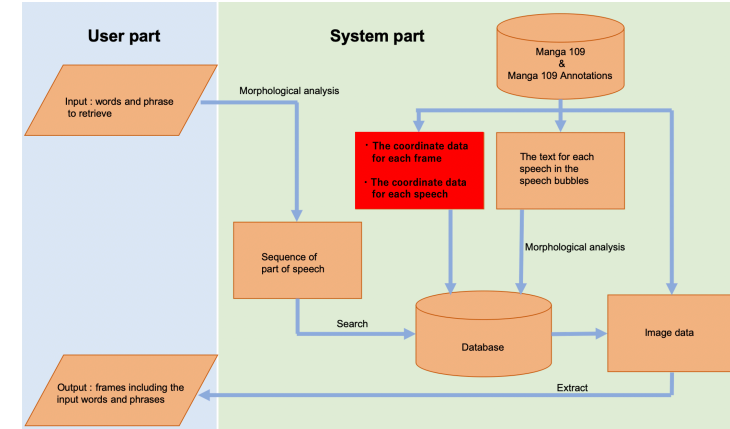
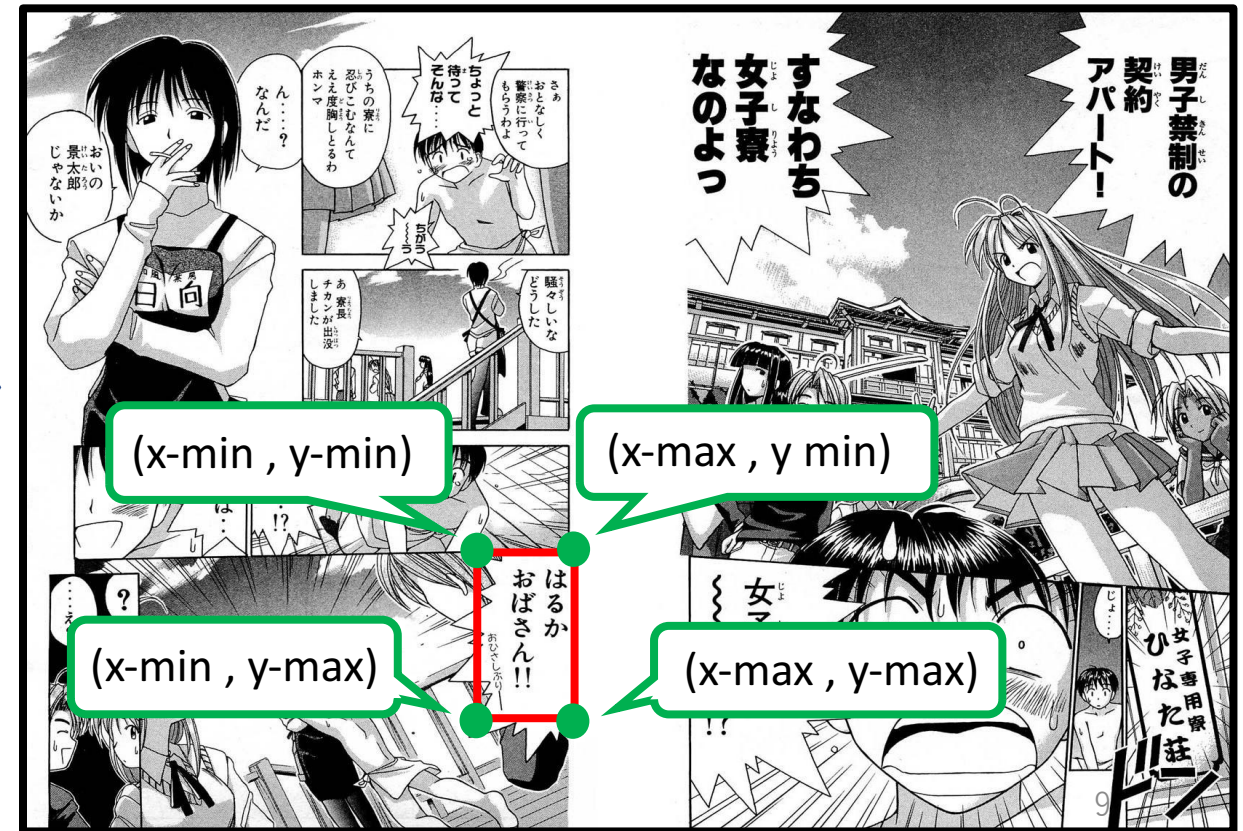
The coordinate data of the speech bubble can be calculated by using **x-min**, **y-min**, **x-max**, and **y-max**.

- The upper-left coordinate is **x-min** and **y-min**
- The upper-right coordinate is **x-max** and **y-min**
- The lower-left coordinate is **x-min** and **y-max**
- The lower-right coordinate is **x-max** and **y-max**

```

<body id="00036146" xmin="1199" ymin="258" xmax="1653" ymax="836" ch
<face id="00036147" xmi
<face id="00036148" xmi
<face id="00036149" xmi
<face id="0003614a" xmi
<body id="0003614b" xmin="20" ymin="29" xmax="344" ymax="618" ch
<text id="0003614c" xmin="629" ymin="794" xmax="744" ymax="1003">は
<text id="0003614d" xmin="913" ymin="60" xmax="1093" ymax="267">すな
<frame id="0003614e" xmin="908" ymin="780" xmax="1449" ymax="1167"/>
<body id="0003614f" xmin="314" ymin="872" xmax="486" ymax="1169" cha
<face id="00036150" xmin="1054" ymin="545" xmax="1112" ymax="603" ch
<text id="00036151" xmin="649" ymin="95" xmax="737" ymax="199">さあお
<text id="00036152" xmin="931" ymin="805" xmax="1046" ymax="1038">女
<body id="00036153" xmin="156" ymin="832" xmax="339" ymax="1169" cha
<text id="00036154" xmin="685" ymin="374" xmax="740" ymax="472">騒々
<frame id="00036155" xmin="904" ymin="2" xmax="1651" ymax="832"/>
<text id="00036156" xmin="1457" ymin="64" xmax="1619" ymax="271">男
  
```

The coordinate data of the speech bubble



The proposed system process the obtained annotations (2)

The proposed system processes the coordinate data of the frame.

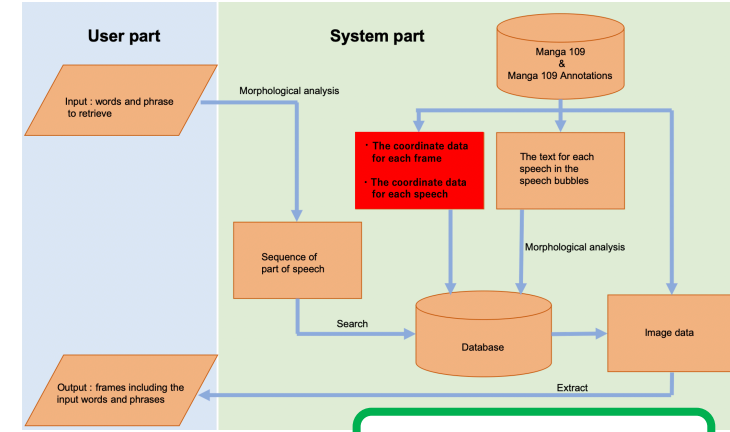
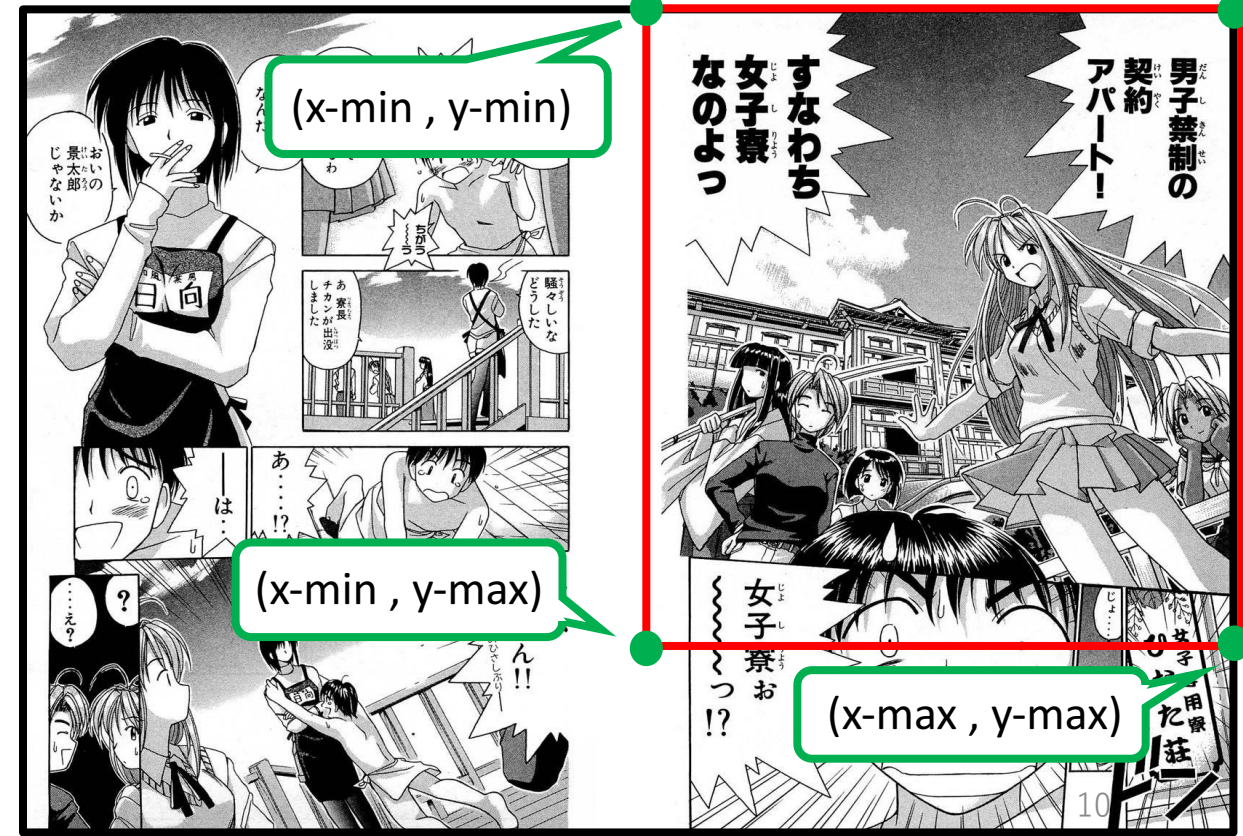
The coordinate data of the frame can be calculated by using **x-min**, **y-min**, **x-max**, and **y-max**.

- The upper-left coordinate is **x-min** and **y-min**
- The upper-right coordinate is **x-max** and **y-min**
- The lower-left coordinate is **x-min** and **y-max**
- The lower-right coordinate is **x-max** and **y-max**

```

<body id="00036146" xmin="1199" ymin="258" xmax="1653" ymax="836" ch
<face id="00036147" xmin="35" ymin="1011" xmax="70" ymax="1090" cha
<face id="00036148" xmin="1139" ymin="651" xmax="1191" ymax="698" ch
<face id="00036149" xmin="68" ymin="620" xmax="203" ymax="776" ch
<face id="0003614a" xmin="1456" ymin="961" xmax="1493" ymax="
<body id="0003614b" xmin="20" ymin="29" xmax="344" ymax="61
<text id="0003614c" xmin="629" ymin="794" xmax="744" ymax="
<text id="0003614d" xmin="913" ymin="60" xmax="1093" ymax=
<frame id="0003614e" xmin="908" ymin="780" xmax="1449" ymax="1167"/>
<body id="0003614f" xmin="
"1169" cha
<face id="00036150" xmin="
ax="603" ch
<text id="00036151" xmin="
"199">さあお
<text id="00036152" xmin="
ymax="1038">女
<body id="00036153" xmin="156" ymin="832" xmax="339" ymax="1169" cha
<text id="00036154" xmin="685" ymin="374" xmax="740" ymax="472">騒々
<frame id="00036155" xmin="904" ymin="2" xmax="1651" ymax="832"/>
<text id="00036156" xmin="1457" ymin="64" xmax="1619" ymax="271">男
  
```

The coordinate data of the frame



(x-max , y min)

(x-max , y-max)

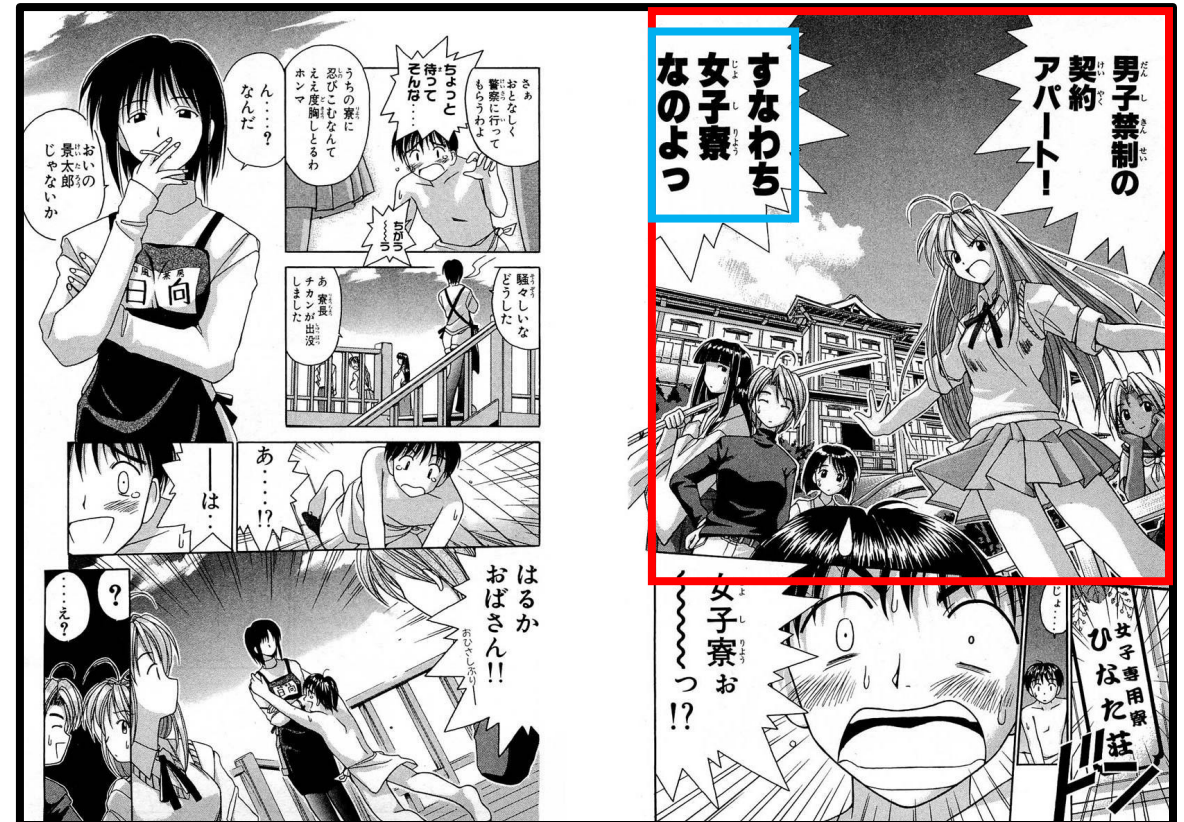
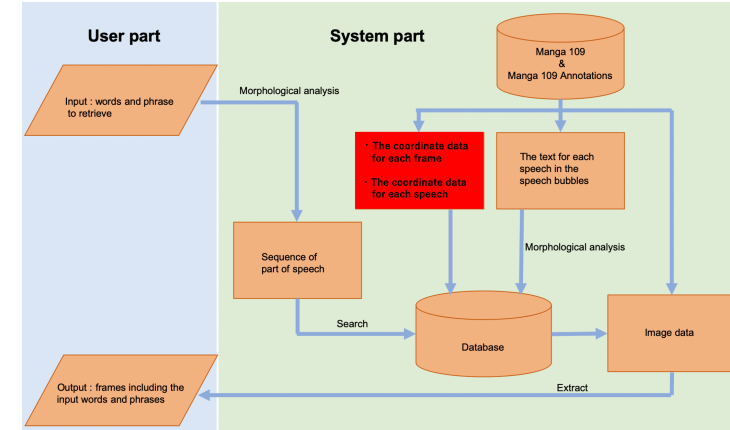
The proposed system process the obtained annotations (3)

The proposed system process the speech bubble and frame.

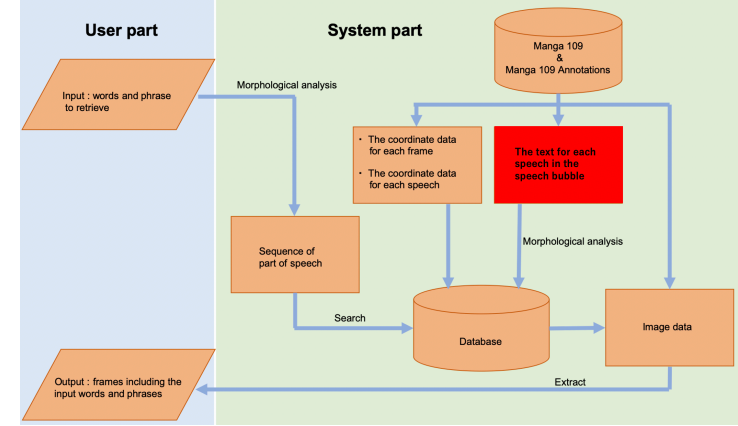
- The red frame is the coordinate data of the frame
- The blue frame is the coordinate data of the speech bubble

The speech in the blue rectangle is included in the frame of the blue rectangle.

→The proposed system detects the frame including the speech.



The proposed system process the obtained annotations (4)



The proposed system process the text of the speech for the speech bubble.

The text of the speech is morphologically analyzed.

The proposed system used MeCab for morphological analysis and NEologd for the dictionary.

```

653" ymax="836" character="00035f83"/>
" ymax="1090" character="00035f83"/>
191" ymax="698" character="00035f83"/>
" ymax="776" character="00035f83"/>
493" ymax="998" character="00035f83"/>
" ymax="618" character="00036141"/>
4" ymax="1003">はるかおばさん！！ おひさしぶりー</text>
3" ymax="267">すなわち女子寮なのよっ</text>
449" ymax="1167"/>
6" ymax="1169" character="00036141"/>
112" ymax="603" character="00035faa"/>
" ymax="199">さあおとなしく警察に行ってもらわよ</text>
46" ymax="1038">女子寮お~~~~っ！？</text>
9" ymax="1169" character="00035f83"/>
0" ymax="472">騒々しいなどうした</text>
1" ymax="832"/>
19" ymax="271">男子禁制の契約アパート！</text>
    
```

The text of the speech for the speech bubble



はるかおばさん！おひさしぶりー
(Aunt Haruka! Long time no see.)



はるか / おばさん / ！
noun noun symbol

お / ひさし / ぶり / ー
prefix noun noun symbol

The proposed system process the input query from the user

The proposed system process the input query.

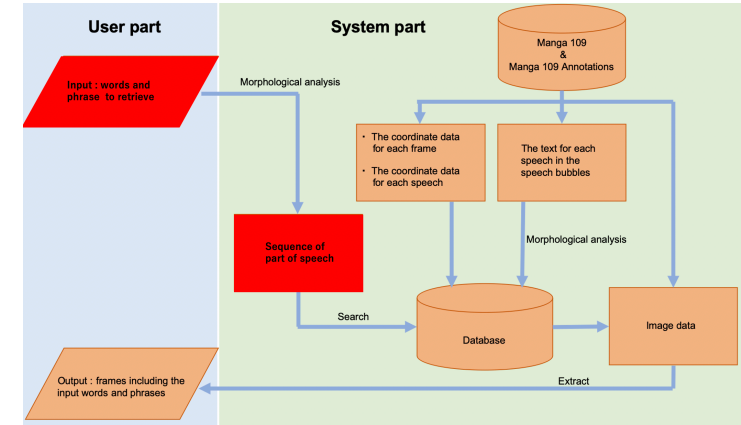
The proposed system accepts the query from the user, i.e., the teacher for the Japanese teacher.

The input queries are three types.

The query

- words
- phrases
- context

As same as obtained annotations, the input query is morphologically analyzed.

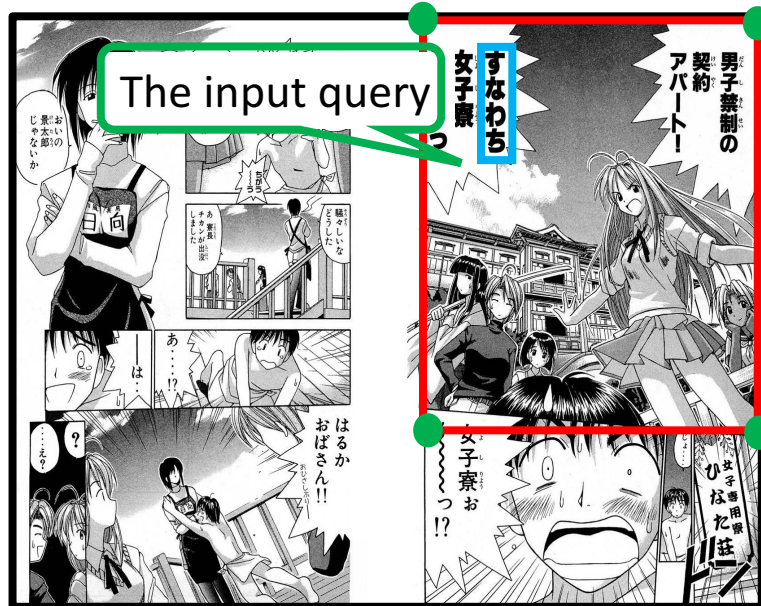
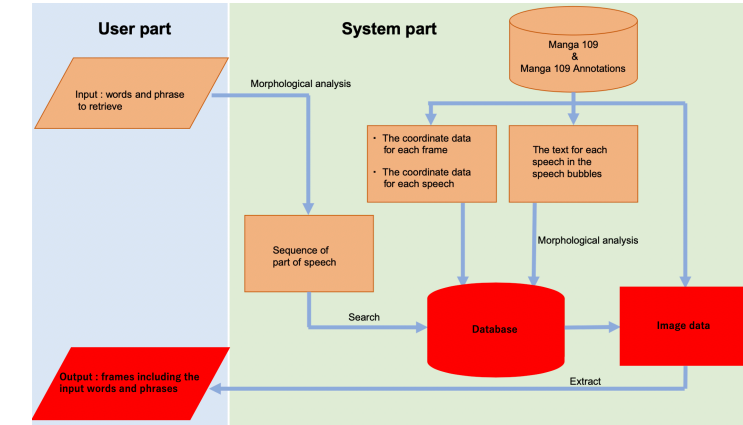


The proposed system outputs the frames including the query

The proposed system process the Manga frame.

The proposed system searches which area on the page of Manga includes the sequence of words with the part-of-speech pf the input query.

→ The proposed system crops the coordinate of each frame including the input query from the images of the spread two pages.



User test settings

We conducted two types of user tests to verify the retrieval results and the effectiveness of the retrieved frames as teaching materials.

We used 10 comics from Manga 109 dataset in two types of user tests.

Title	Author
Belmondo	Shoei Ishioka
EverydayOsakanaChan	Yuka Kuniki
HarukaRefrain	Shinpei Ito
JijiBabaFight	Shinji Nishikawa
LoveHina	Ken Akamatsu
NichijouSoup	Uni Shindou
OL_Lunch	Youko Sanri
UchiNoNyan'sDiary	Gasan
UnbalanceTokyo	Minako Uchida
YamatoNoHane	Kaori Saki

User test(The retrieval of Manga frames)

We used 10 kinds of adverbs and conjunctions for the input queries in this user test.

We have several reasons for choosing adverbs and conjunction for input queries.

- Adverbs and conjunctions are complicated to use in Japanese language education so those are needed to learn the context to be use.
- Adverbs and conjunctions are independent words and have no conjugation in the Japanese language.

Adverb	Conjunction
ますます	それとも
いまにも	だけど
おそらく	つまり
すでに	ところが
せっかく	ところで

User test results(The retrieval of Manga frames) (1)

We show the statistical result of the retrieval with each word, which is set as the retrieval target i.e., adverbs and conjunctions.

A total of **66 frames** were detected as the frames including each from a total of 10 comics.

	ますます	いまにも	おそらく	すでに	せっかく	それとも	だけど	つまり	ところが	ところで
Belmondo	0	0	0	0	0	0	1	5	1	1
EverydayOsakanaChan	2	0	0	3	3	0	1	0	1	4
HarukaRefrain	0	0	0	0	2	0	0	2	0	0
JijiBabaFight	1	0	1	2	1	0	0	0	0	1
LoveHina	0	0	0	0	1	0	0	0	0	0
NichijouSoup	0	0	0	0	2	0	0	0	0	2
OL_Lunch	0	0	0	1	1	0	0	1	0	1
UchiNoNyan'sDiary	0	0	2	0	3	1	0	0	4	0
UnbalanceTokyo	0	0	0	0	1	2	1	4	1	3
YamatoNoHane	0	0	0	0	0	0	1	2	0	0

User test results(The retrieval of Manga frames) (2)

In the result, **17 frames** out of 66 frames showed the conversation between more than two characters. Tough it is not so much than we expected, the proposed system automatically found the frames showing the conversation using the intended words.

Examples of frames



JijiBabaFight © Shinji Nishikawa



UnbalanceTokyo © Minako Uchida



HarukaRefrain © Shinpei Ito

User test(The evaluation of the retrieved frames)

1. We prepared three participants as the test user.
2. They were university students whose led is computer science. And they had no experience of language instruction before.
3. They chose adverbs and conjunctions that should be necessary for learning Japanese from the textbook “Genki”.
4. They used the proposed system to retrieve frames used as the sub materials from three comics.
5. Takei’s teacher subjectively evaluated the retrieved results.

User test results (The evaluation of the retrieved frames) (1)

We show the statistical results of the evaluation.

	あまり	でも	まだ	もう	ちょっと	よく	ところで
Good	33.3%	58.6%	16.7%	29.4%	61.5%	16.7%	100%
Bad	66.7%	41.4%	83.3%	70.6%	38.5%	83.3%	0%

The retrieved results for “ところで”, which is a conjunction, were good as the teaching materials.

The retrieved frame showed a conversational context with only one.

Examples of frames



NichijouSoup © Uni Shindou



NichijouSoup © Uni Shindou

User test results (The evaluation of the retrieved frames) (2)

We show the statistical results of the evaluation.

	あまり	でも	まだ	もう	ちょっと	よく	ところで
Good	33.3%	58.6%	16.7%	29.4%	61.5%	16.7%	100%
Bad	66.7%	41.4%	83.3%	70.6%	38.5%	83.3%	0%

The retrieved results for “あまり”, which is an adverb, not good. The word has several types of usages and is used in different contexts. So, some of the retrieved frames were evaluated as inappropriate for teaching materials.

Examples of frames



EverydayOsakanaChan © Yuka Kuniki



EverydayOsakanaChan © Yuka Kuniki

Conclusions

- The proposed system reduced the time for retrieval than human manual retrieval.
- From the user test, it was confirmed that the participants retrieved the frames appropriate teaching materials for some of the words.

In the future

- The conversation may not be in one frame but be continued over multiple frames.
 - The multiple frames should be provided as the retrieved results.
- The retrieval should be controlled for usage and meaning of the word.