Extracting Tip Information from Social Media

Konan University ☆Yuki Hattori

Akiyo Nadamoto

Agenda

1. Back ground

2. Definition of tip information

3. Technical points
   - Extracting tip information
     – Experience mining, Tip keywords
   - Calculating ranking

4. Prototype system, Experiments

5. Conclusion, Future work

Background 1

There are many kinds of social media on the internet.

They have created numerous and diverse community.

By using social media

Easy for users to post and exchange information.

personal behavior, experience, their own sentiments

This information is not written in ordinary web pages.

Background 2

Some festival communities in a social networking services(SNS)

Presenting experience information that is not written on official web pages.

"When you go to the Suwa Fireworks Festival by car, you should exit the expressway one exit early. If you exit at the nearest exit, you will hit a terrible traffic jam."

Community

important   important

inside users   outside users
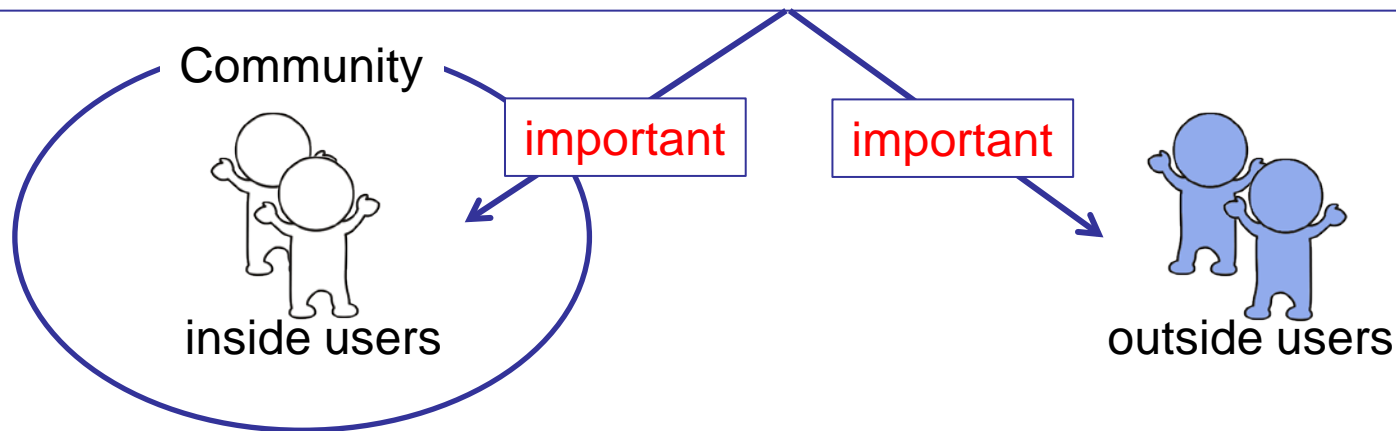
That happens to be important information for users who are not only community members, but also for people who are outside the community.

Background 3

Difficult to extract **important information** from social media.

➡️ So much information exists on social media.

"*Tip information*"

We propose a method to extract tip information from SNSs of social media

The definition of tip information

There are 4 points of the definition of tip information.

> ***The information is credible.***

> ***The information is important.***

> ***A user does not know the information.***

> ***The information is not generally known.***

We target on credible and important information as a first step in extracting the tip information from SNSs.

Definition of tip information

➢ *The information is credible.*

Regarding sentences that are written based on the author's actual experience as credible.

➢ *The information is important.*

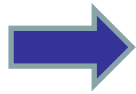The important information uses some ***common keywords***.

"recommend", "should", "hot spot".

"Tip keywords"

→ Extracting the important information which is written using some tip keywords.

Definition of tip information
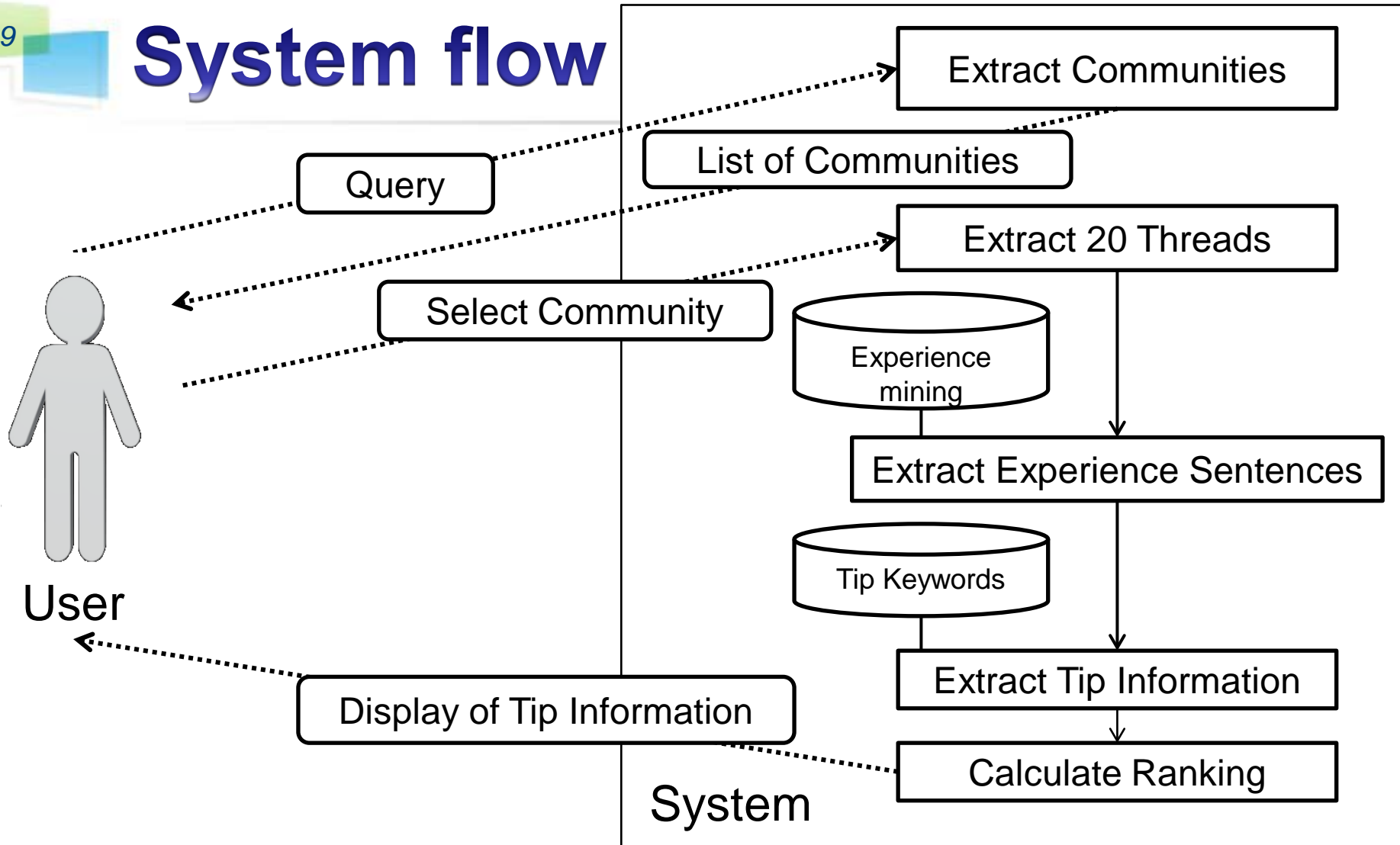
Much tip information in SNSs.

➡️ Difficult for users to extract it.
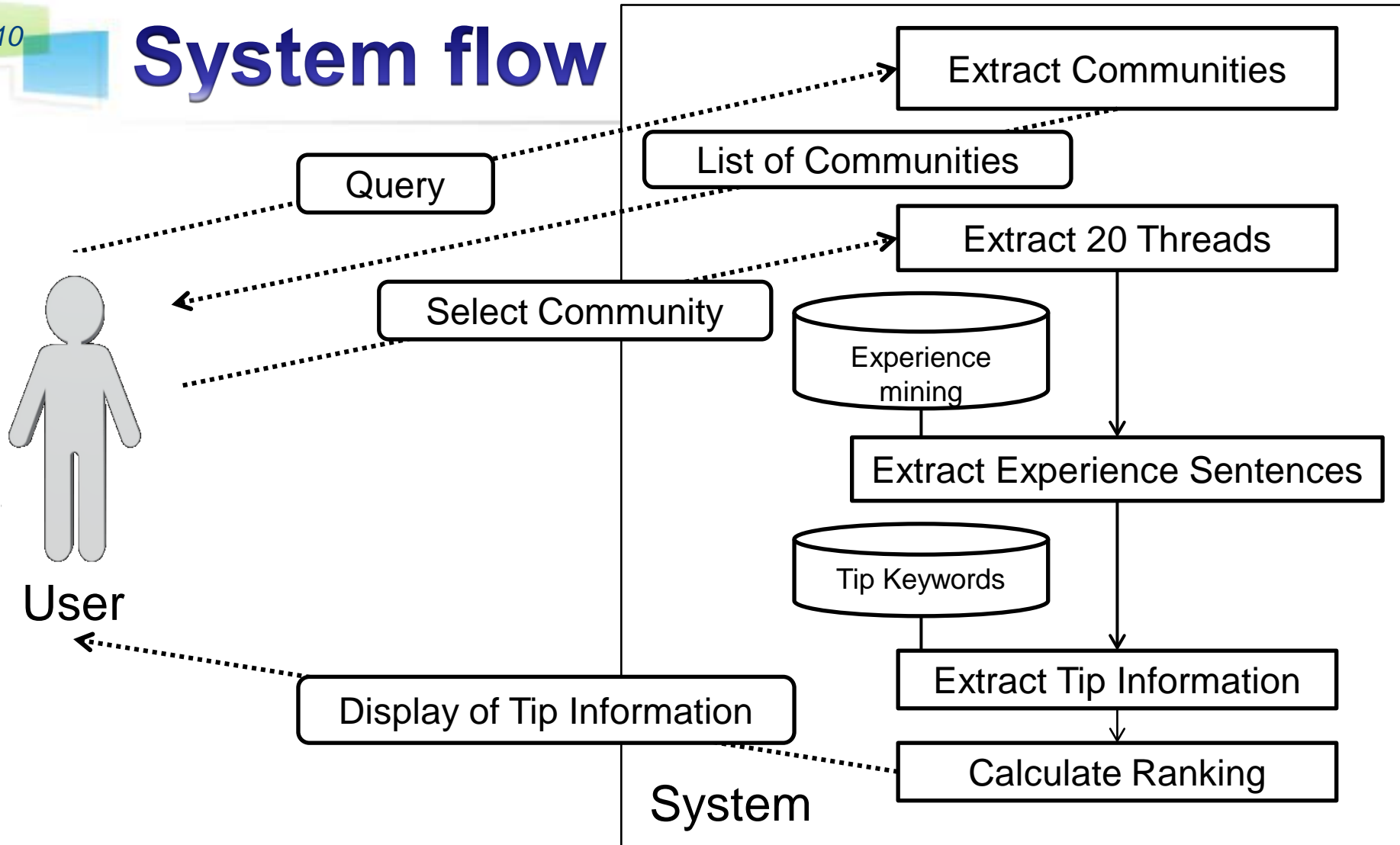
⬇️

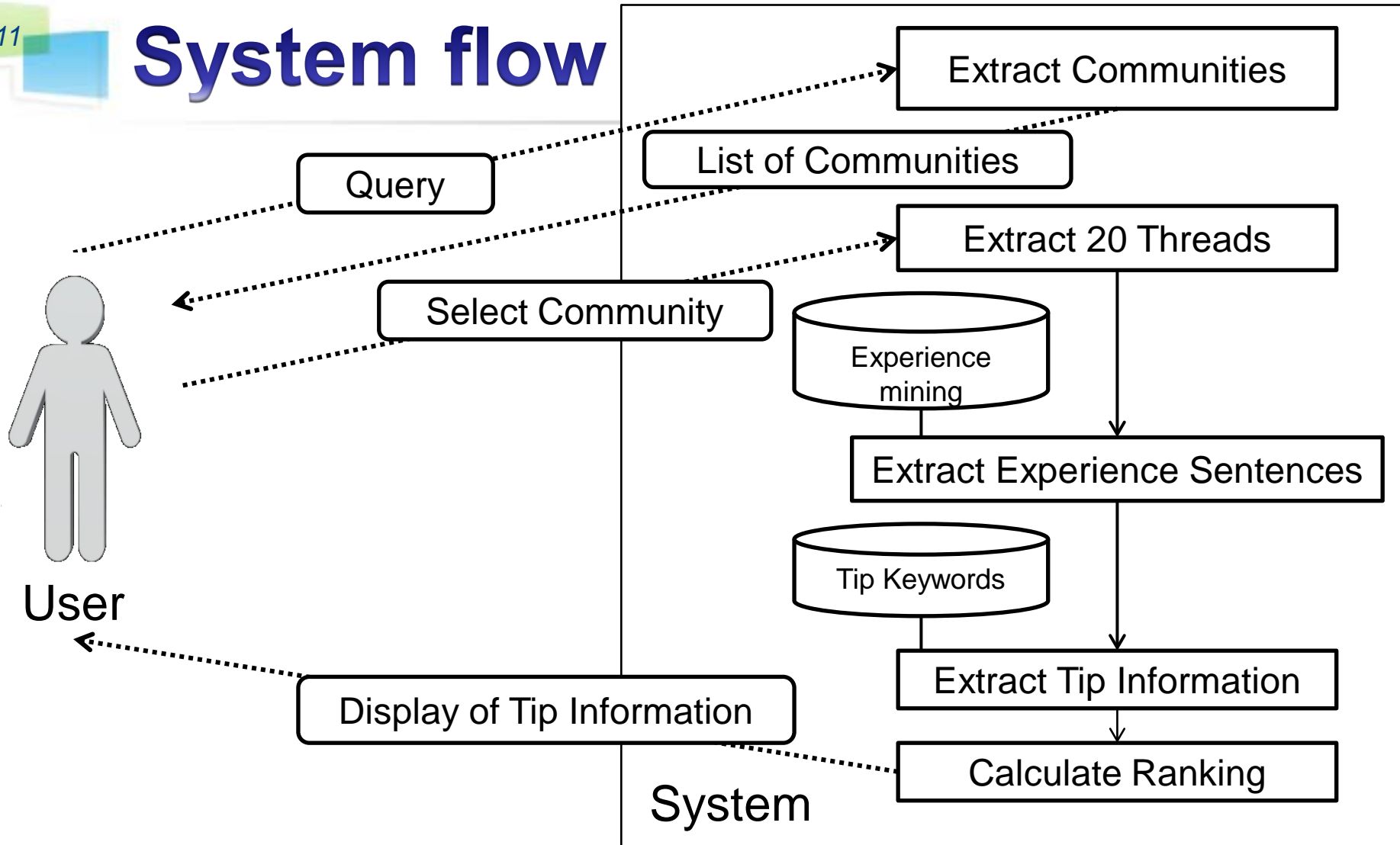Considering *ranking* of the tip information.

System flow

Extract Communities

List of Communities

Query

Extract 20 Threads

Select Community

Experience mining

Extract Experience Sentences

Tip Keywords

User

Display of Tip Information

Extract Tip Information

Calculate Ranking

System

1. The user inputs a query.
2. The system extracts communities from SNS and browses the list of communities.

System flow

Extract Communities

List of Communities

Query

Extract 20 Threads

Select Community

Experience mining

Extract Experience Sentences

Tip Keywords

User

Extract Tip Information

Display of Tip Information

Calculate Ranking

System

3. The user selects a community from the list.
4. The system extracts comments from the community.
5. It extracts actual experience sentences.

System flow

Extract Communities

Query

List of Communities

Extract 20 Threads

Select Community

Experience mining

Extract Experience Sentences

Tip Keywords

Extract Tip Information

User

Display of Tip Information

Calculate Ranking

System

6. It extracts tip information from the actual experience sentences using a tip keyword dictionary.

7. It calculates the ranking degree of the tip information.

8. It browses the tip information based on the ranking.

Extracting tip information

How to extract tip information.

credible   important

One of a credible information is based on the author's experience.

Important information uses some tip keywords.

We extract experience sentences and tip keywords from comments.
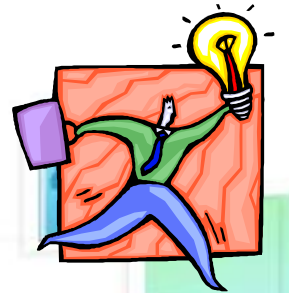
Extracting experience sentences

Extracting experience sentences based on *experience mining*, which was proposed by Inui.

K. Inui, S. Abe, H. Morita, M. Eguchi, A. Sumida, C. Sao, K. Hara, K. Murakami, and S. Matsuyoshi. Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In Proceedins of 49the 2008 IEEE/WIC/ACM International Conference on Web Intelligence, pages 314321, 2008.

Experience mining is intended for the automatic collection of instances of personal experiences from social media.
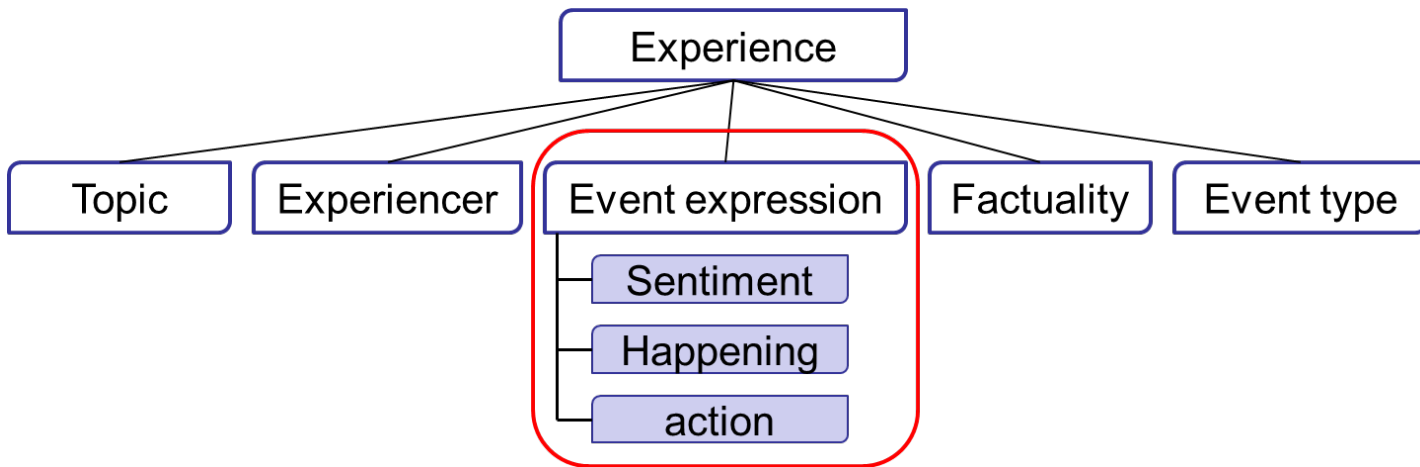
We assume that tip information is based on the author's event expression.

Extracting experience sentences

Using just only sentiment and happening from experience mining dictionary.

Event expression
- Sentiment
- Happening
- action

Experience mining
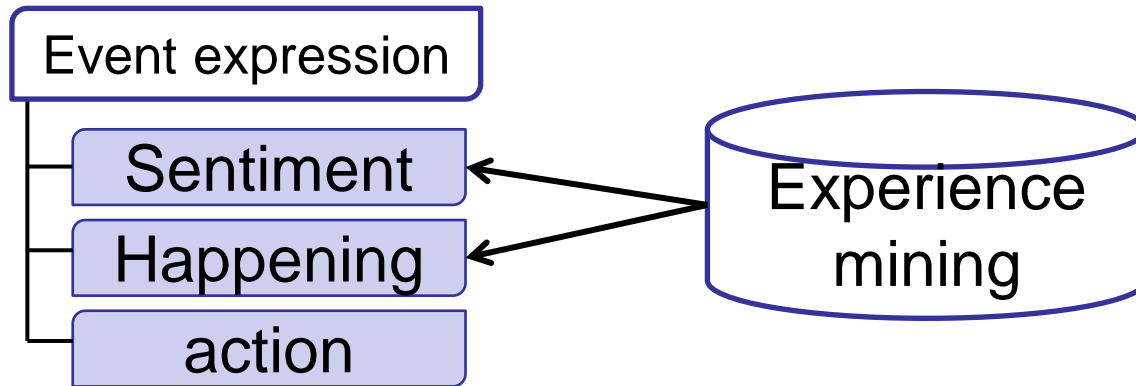
We propose new action words.

⇨ Action words depends on domain.

⇩

We create action words dictionary.

Creating an action words dictionary

Festival domain

Verb		Noun	
Go	Buy	Use	Participate
Lose away	View	Move	Excitement
Able to	Drink	Cheers	Activity

1. Gathering 3000 comments from festival communities.

2. Extracting the action words manually and count the term frequency(TF) of the verbs and nouns.

3. Inferring the top 50 words of TF as festival domain action words.

Extracting experience sentences using an experience mining dictionary and an action words dictionary.

The sentences become tip information candidates.

Extracting tip information

We propose a means of extracting tip information.

credible   important

Considering that credible information is based on the author's experience.

Considering that important information uses some tip keywords.

We propose a means to extract experience sentences and tip keywords.

Creating tip keywords dictionary

The important information uses some ***common keywords***.

"recommend", "should".

"Tip keywords"

Extracting tip keywords using our experiment.

Creating tip keywords dictionary

Experiment flow | Target ➡ Festival communities

1. 5 participants read 2000 comments and judged the comments as tip information or not.

2. Regarding information that is judged as tip information from 4 participants in 5 participants as tip information.

3. Extracting tip keywords from the sentences that were judged as tip information.

Example of tip keywords		
recommendation	.. Is better	How about ..?
should	By all means	Highly recommended

Extracting important tip information by using the tip keyword dictionary.

Ranking tip information

Communities of SNS contain much tip information.

➡️ Difficult for users to know tip information immediately.

⬇️

We consider calculation of the ranking degree of the tip information, and present them by ranking.

Extracting tip keywords

There are some types of tip keyword, and their importance is different.

➡️ 1. Classifying tip keywords into the types.
2. Calculating the ranking using the types that we divide.

Classifying the tip keywords

We classify the tip keywords as four types

A recommendation type

Some comments recommend to the user something with which the author has had an experience.

A catchphrase type

The word that is used as a catchphrase of the advertisement makes the user feel a profitable feeling.

An impression type

An impression of an author who has had actual experience is important for users.

An emphasis type

An emphatic comments are more important than other comments.

A recommendation type
Recommendation
.. Is better
A catchphrase type
Free
Low price
An impression type
Love at first sight
Happy
An emphasis type
dramatic
Pretty good

Classifying the tip keywords manually into four types.

Ranking

We classify tip keywords of four types.

➡ Multiple types exist among comments and impressive tip keywords differ according to the type.

We examined four parties to ascertain which type was the most impressive.

Impression degree

| A recommendation type |
| A catchphrase type |
| An impression type |
| An emphasis type |
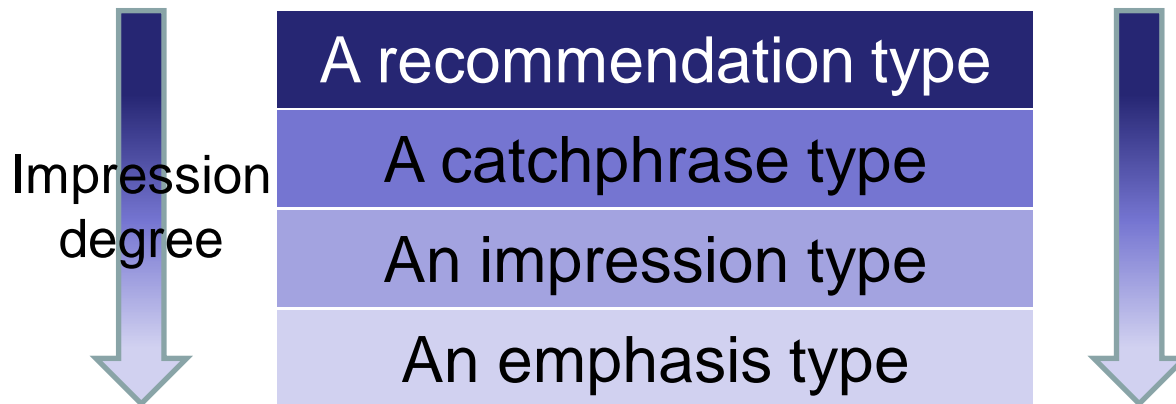
The recommendation type is inferred as the most impressive of tip keywords.

Ranking

Impression degree

A recommendation type
A catchphrase type
An impression type
An emphasis type
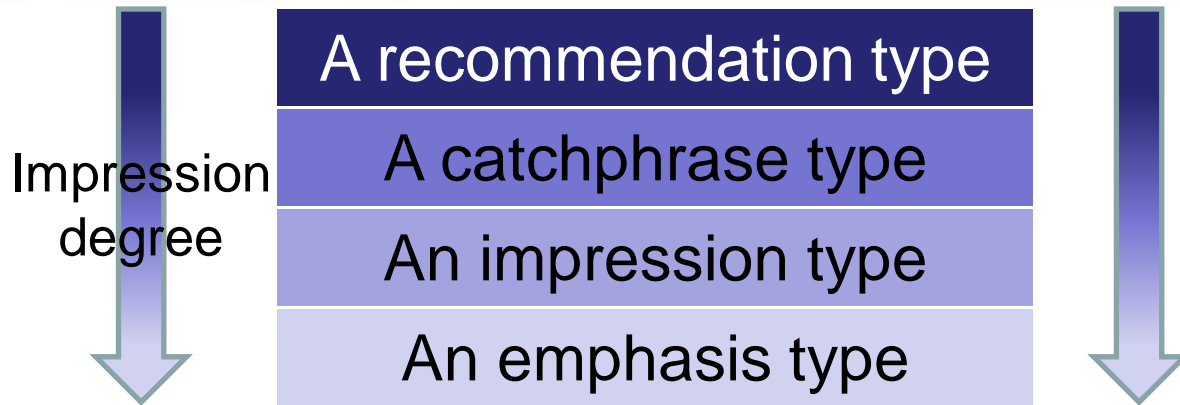
➢ Define the ranking expression.

$$M = \alpha RE + \beta CA + \gamma IM + \delta EM$$

$$(\alpha > \beta > \gamma > \delta)$$

M: Ranking degree
RE: Recommendation type
CA: Catchphrase type
IM: Impression type
EM: Emphasis type
α, β, γ and δ: parameters

We calculate ranking by this equation.

Prototype System

We developed a prototype system using our proposed method.

The results display

Experiments

We had two experiments.

Experiment 1: Availability of extracting tip information.

Experiment 2: We measure the weight of expression of ranking.($M = \alpha RE + \beta CA + \gamma IM + \delta EM$).

Experiment 1

➢ Conducting an experiment of availability of tip information extraction.

- The datasets are 9 themes of communities that discuss festival.
- 5 participants judged the comments.

Community	Number of all comments	Number of comments include tip information	Average of precision
PL Fireworks Show	909	114	63%
Koyabu Sonic	891	28	31%
Nabana no Sato	262	30	55%
Beach in Kansai area	318	16	72%
Kyoto Gion Festival	441	29	49%
Autumn Leaves in Kyoto	329	36	66%
Kobe Luminarie	396	18	23%
Lake Suwa Fireworks Show	1313	128	48%
Gathering of clams	415	24	63%
--	--	--	52%

Experiment 1 : Discussion

In a good case

PL Fireworks show

I denitely recommend taking a train. The train on the way home is super-crowded, so I advise you to get on at Tondabayashi-nishiguchi Stn(the previous Stn). or Kawanishi Stn(next Stn)., even if it means walking one station further. Driving a car makes you miserable. My relatives living in Seika Town, Kyoto once came by car. They say that they left Tondabayashi at 11 pm and got home at 4 am.
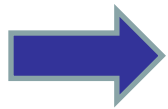
Experiment 1 : Discussion

In a bad case

Many comments discuss another theme in the communities.

Kobe Luminarie which is a name of christmas illumination:
Few people discuss Kobe Luminarie itself, but they discuss another night view in Kobe.

➡️ Kobe is famous for it beautiful night view.

⬇️

We should consider that the discuss theme is fit to the community theme.

Experiment 2

We had measured the weight of expression of ranking.

We set the appropriate weight α, β, γ, δ and threshold ranking degree in expression by changing them.

$$M = αRE + βCA + γIM + δEM$$
$$(α > β > γ > δ)$$

M: Ranking degree
RE: Recommendation type
CA: Catchphrase type
IM: Impression type
EM: Emphasis type
α, β, γ and δ: parameters

The datasets are 4 types: "PL Fireworks show", "Beach in Kansai area", "Autumn Leaves in Kyoto", "Kyoto Gion Festival".
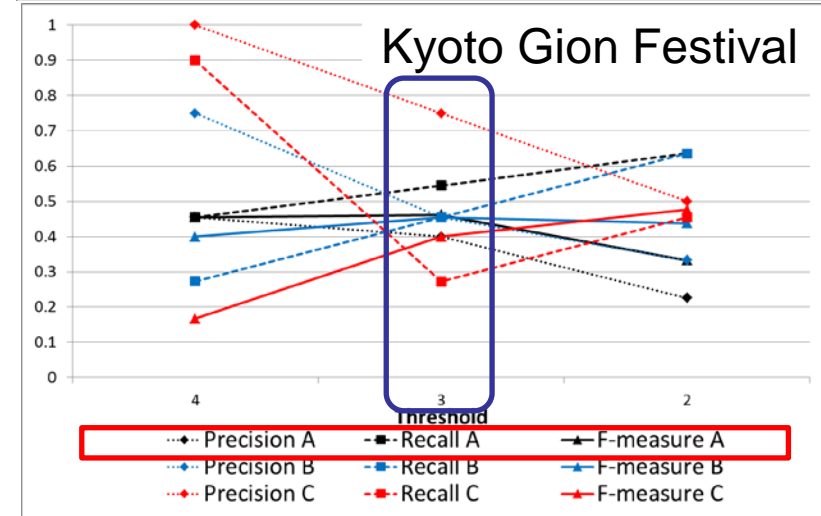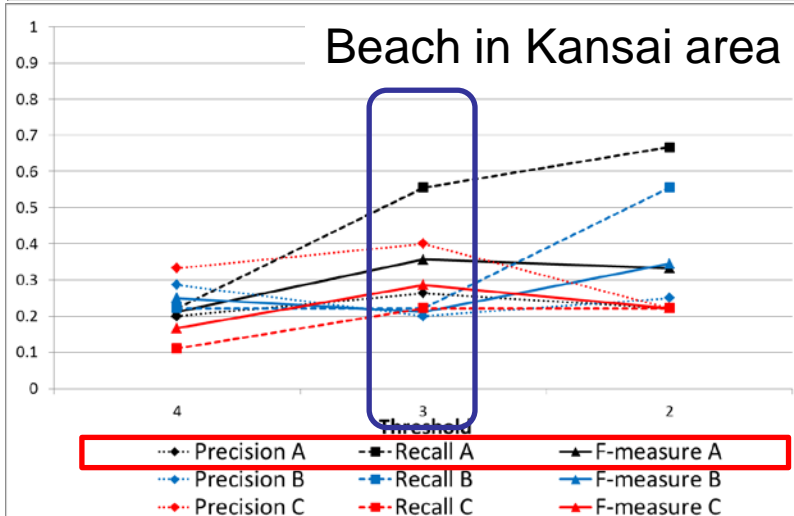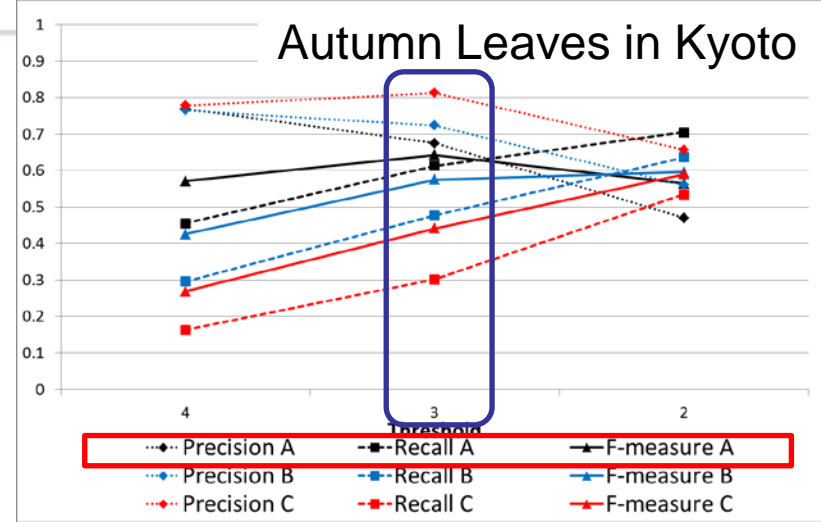
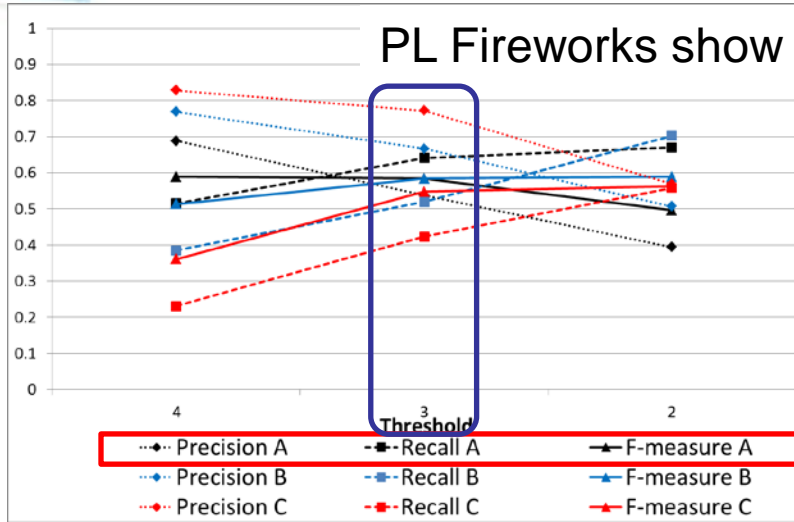Set thresholds of 4, 3, and 2 for each dataset.

The value of α, β, γ, and δ in all results are like this.

A type : α=1.0, β=0.9, γ=0.8, and δ=0.7
B type : α=1.0, β=0.8, γ=0.7, and δ=0.3
C type : α=1.0, β=0.7, γ=0.4, and δ=0.1

Regarding α as the basis. Its value is 1.0.

Results of Experiment 2

PL Fireworks show

Autumn Leaves in Kyoto

Beach in Kansai area

Kyoto Gion Festival

Determining the α as 1.0, β as 0.9, γ as 0.8, δ as 0.7, and the threshold as 3.0.

The average precision is 54.4%, the average recall is 58.9%, and the average F-measure is 0.52

Conclusion

We proposed a method for extracting tip information that is credible and important information from SNS.

1. We proposed how to extract tip information from SNS.

2. We proposed how to rank the tip information.

3. We developed a prototype system and experiments.

Future Work

1. We should consider that the content of comments fits the community theme.

2. We do not consider the comment context, but we should.

3. We should consider personalization.
   the information should be regarded as tip information depends on the person.