CAO: A Fully Automatic Emoticon Analysis System

Michal Ptaszynski, Jacek Maciejewski,
Pawel Dybala, Rafal Rzepka and Kenji Araki

Graduate School of Information Science and Technology
Hokkaido University

Presentation Outline

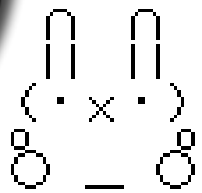
- Emoticons - Definition
- Database Construction
- CAO – Emoticon Analysis System
- Evaluation of CAO
- Conclusions and Future Work

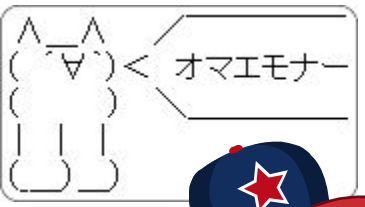
Emoticons - Definition

Our working definition of emoticons…

Emoticons - Definition

Our working definition of emoticons…





☆＊ ヽ(。・ω・)ﾉ Hello ♡(´・ω・`)

(°,_ヽ°) ☆ (#`皿´) Angry

Smile (＊´艸`) (°Д°;)

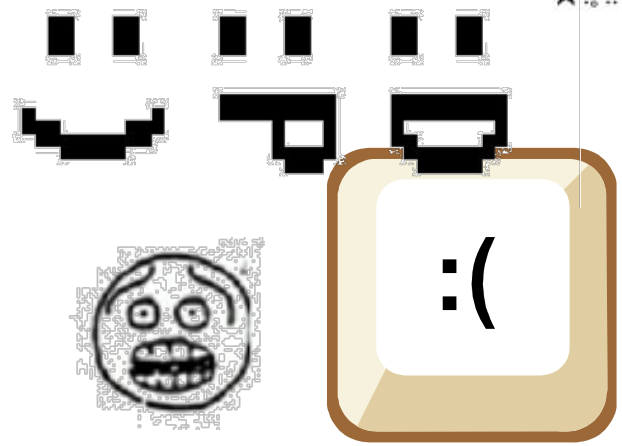
└(・∀・)┐ Keep from laughing out loud

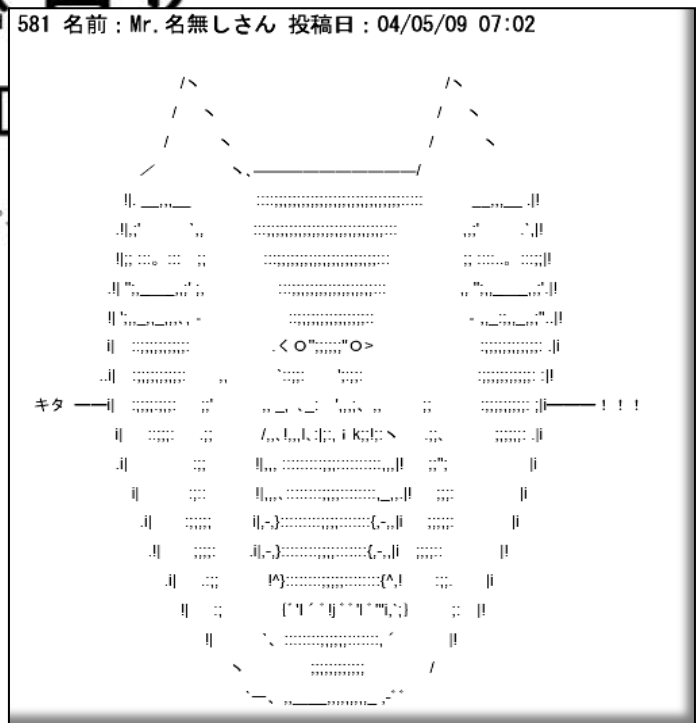
♪ Dance

(^3^) Laugh ＊ Щ(´

kiss ヾ(≧▽≦)ﾉ ☆

☆.。.:*・°☆.。.:*・°☆.。.:*・°☆.。.:*・°☆.。.:*・°



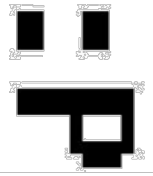


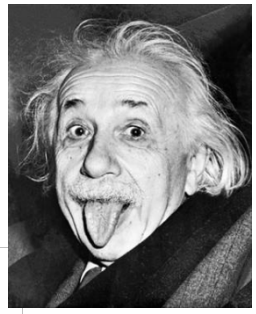
Emoticons - Definition

Emoticons:

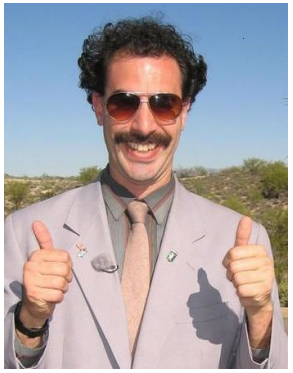
- Emoticons are representations of body language in online communication (more-less).

щ(´□`)щ





(・ × ・)



Therefore…

Emoticons - Definition

Emoticons:

- Are an important in part of communication [1,2] in online communities (blogs, forums, BBS, e-mails, chat-rooms, etc.)

1. Suzuki, N. and Tsuda, K. 2006. Automatic emoticon generation method for web community, WBC2006, pp. 331-334.
2. Derks, D., Bos, A.E.R., von Grumbkow, J. 2007. Emoticons and social interaction on the Internet: the importance of social context, Computers in Human Behavior, 23, pp. 842-849.

Emoticons - Definition

Emoticons:

- But sometimes are difficult to understand

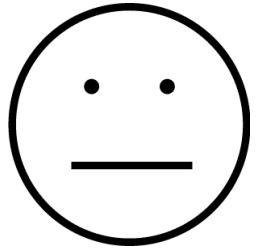
(☣_ ≻ ☣)

7:^)~

(。ᵕ‿ᵕ)

(/~—~)/..:: : * : ☆* °

(^ ·◇· ^)



Emoticons - Definition

Emoticons:

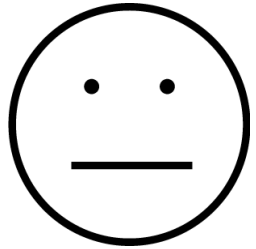
- But somet

Need to analyze them effectively

(°ー°〃)

(/~—~)/..:: : * : ☆* °

(^ ;◇; ^)



Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)

Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)

Emoticons - Definition

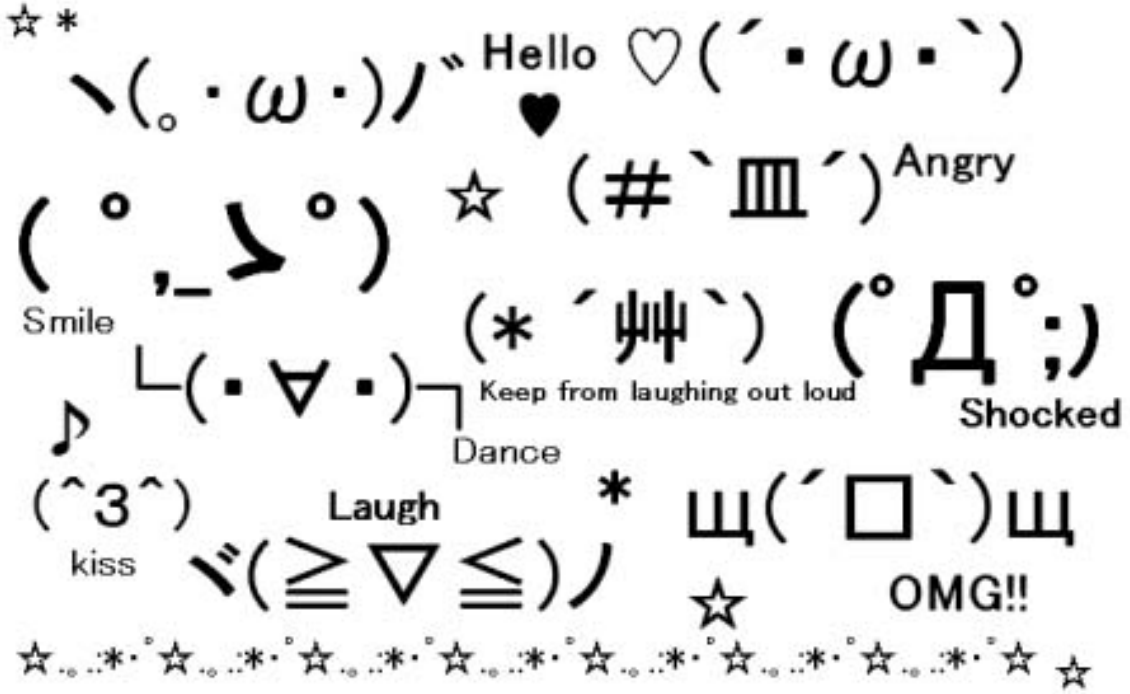
Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)
  - 1-line Eastern

Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)
  - 1-line Eastern



Emoticons - Definition

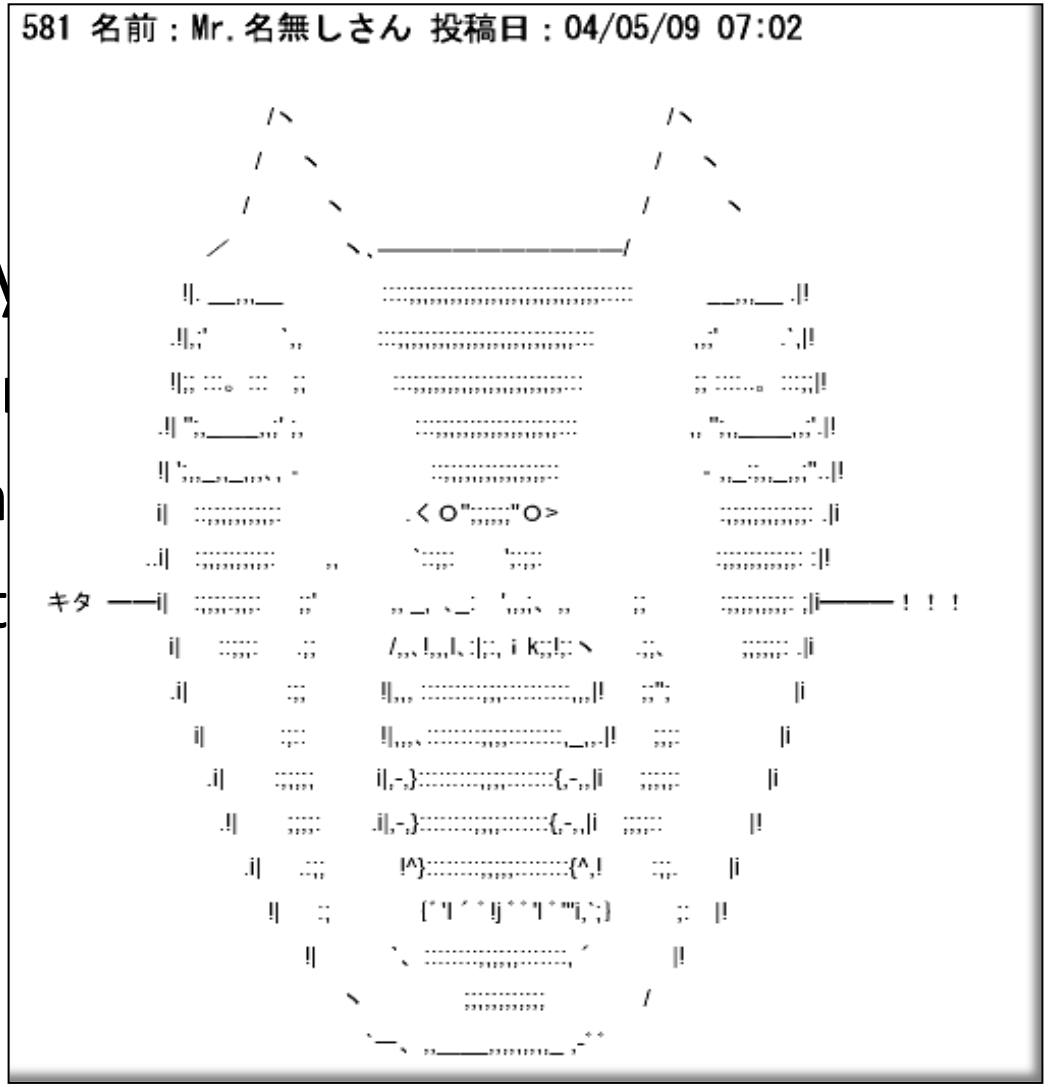
Emoticons:

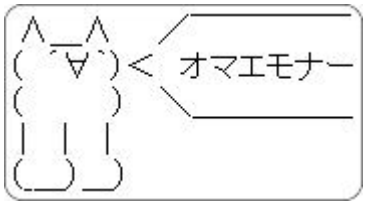
- Can be roughly divided into:
  - 1-line Western (text-base or pictures)
  - 1-line Eastern
  - Multiline Eastern

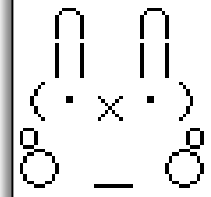
Emoticons - Definition

Emoticons:

- Can be roughly
  - 1-line Western
  - 1-line Eastern
  - Multiline East







Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)
  - 1-line Eastern
  - Multiline Eastern

We focused on these, because…

Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western
  - 1-line Eastern
  - Multiline Eastern

← **There already is some research + we were a little more ambitious**

← **We are not that crazy**

Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western
  - 1-line Eastern
  - Multili[illegible] Eastern

← There already is some research + we were a little more ambitious

← We are not that crazy

Only a little research done here

Emoticons - Definition

Some examples:

\(*^o^*)/

・゜・(ノД`;)・゜・

(;^_^A

(°.　°)

(^_-)y--~

(==╥==Д==╥==)

(｡･_･｡)人(｡･_･｡)

（＊￣ー￣）

Emoticons - Definition

Some examples:

\(*^o^*)/

·゚·(ノд`;)·゚ ·

(;^_^A

(°.　°)

(^_-)y--~

(==╦==д==╦==)

(｡･_･｡)人(｡･_･｡)

（＊￣ー￣）

Suddenly came inspiration!

Emoticons - Definition

Some examples:

\(*^o^*)/

･ﾟ･(ﾉД`;)･ﾟ･

(;^_^A

(°.　°)

(^_-)y--~

(==╥==Д==╥==)

(｡･_･｡)人(｡･_･｡)

(＊￣ー￣)

Suddenly came inspiration!

Since emoticons are representations of body language…

Emoticons - Definition

Some examples:

\(*^o^*)/

Suddenly came

･ﾟ･(ﾉД`;)･ﾟ

(;^_^A

(°.　°)

s are

(^_-)y--~

s of

(==╦==Д==

…

(。･_･。)人(。･_･。)

(*　　　　　)

A structural approach to body language could be applicable here as well!

Emoticons - Definition

- **Theory of kinesics:**
- Non-verbal behavior is used in everyday communication systematically and can be described structurally.
- A minimal part = a kineme, the smallest meaningful set of body movements, e.g. raising eyebrows, etc.

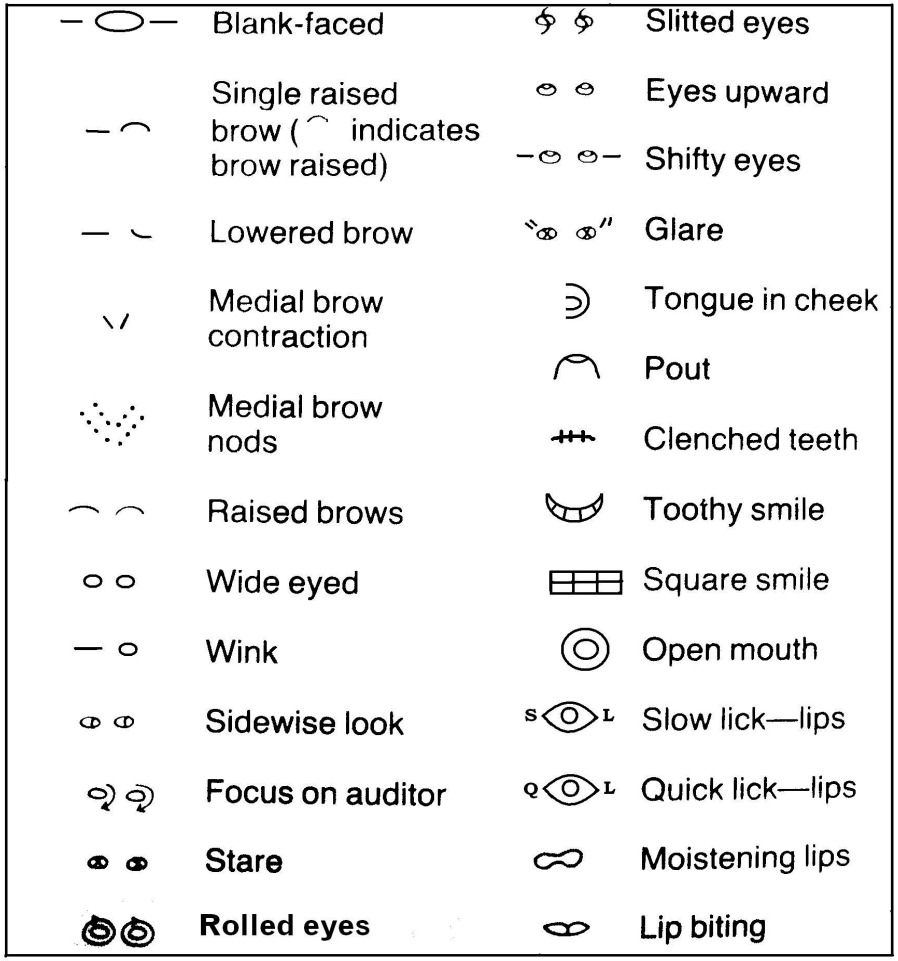
Birdwhistell (1952, 1970)

Birdwhistell, R. L. 1952. Introduction to kinesics: an annotation system for analysis of body motion and gesture, University of Kentucky Press.
Birdwhistell, R. L. 1970. Kinesics and Context, University of Pennsylvania Press, Philadelphia.

Emoticons - Definition

- **Theory of kinesics:**
- Non-verbal behavior is
  communication system
  described structurally.
- A minimal part = a kine
  meaningful set of body
  raising eyebrows, etc.

B



Birdwhistell, R. L. 1952. Introduction to kinesics: a
motion and gesture, University of Kentucky Press.
Birdwhistell, R. L. 1970. Kinesics and Context, University of Pennsylvania Press, Philadelphia.

Emoticons - Definition

Some examples:

\(*^o^*)/

·゜·(ノД`;)·゜·

(;^_^A

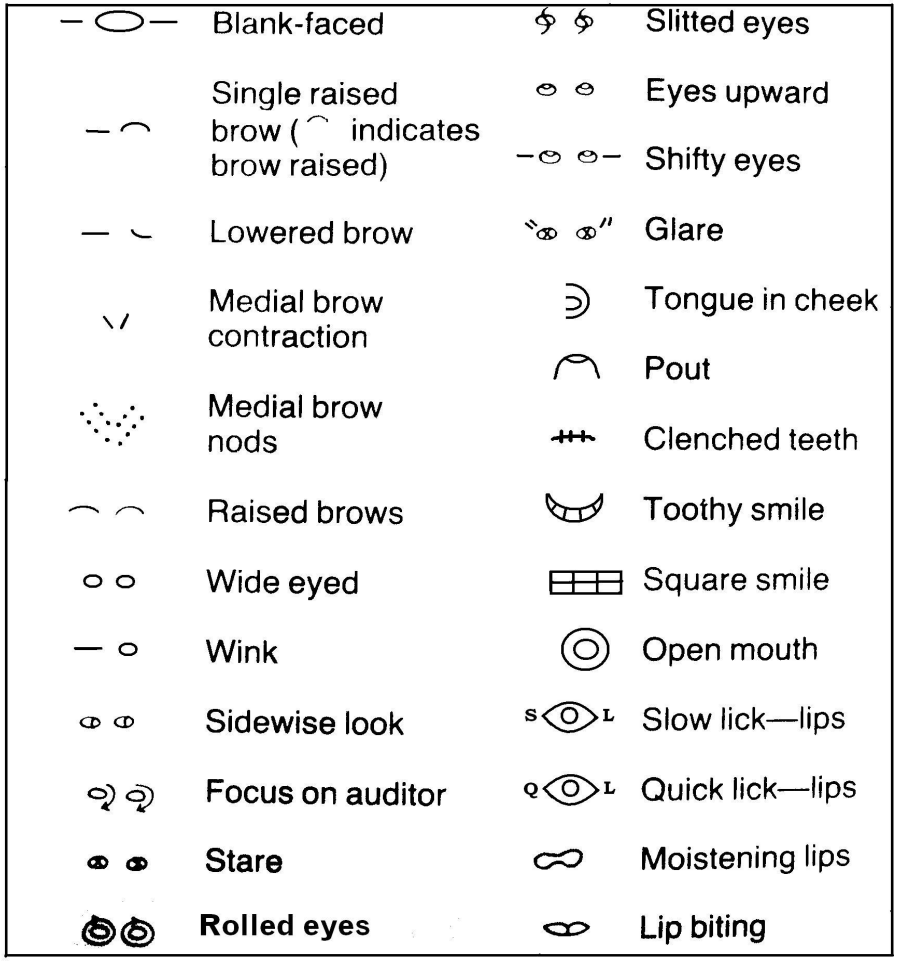
(°.　°)

(^_-)y--~

(==╦==Д==╦==)

(。・_・。)人(。・_・。)

(*　　　　　　　　　　　　　　)



Emoticons - Definition

Some examples:

\(*^o^*)/

・゜・(ノД`;)・゜・

(;^_^A

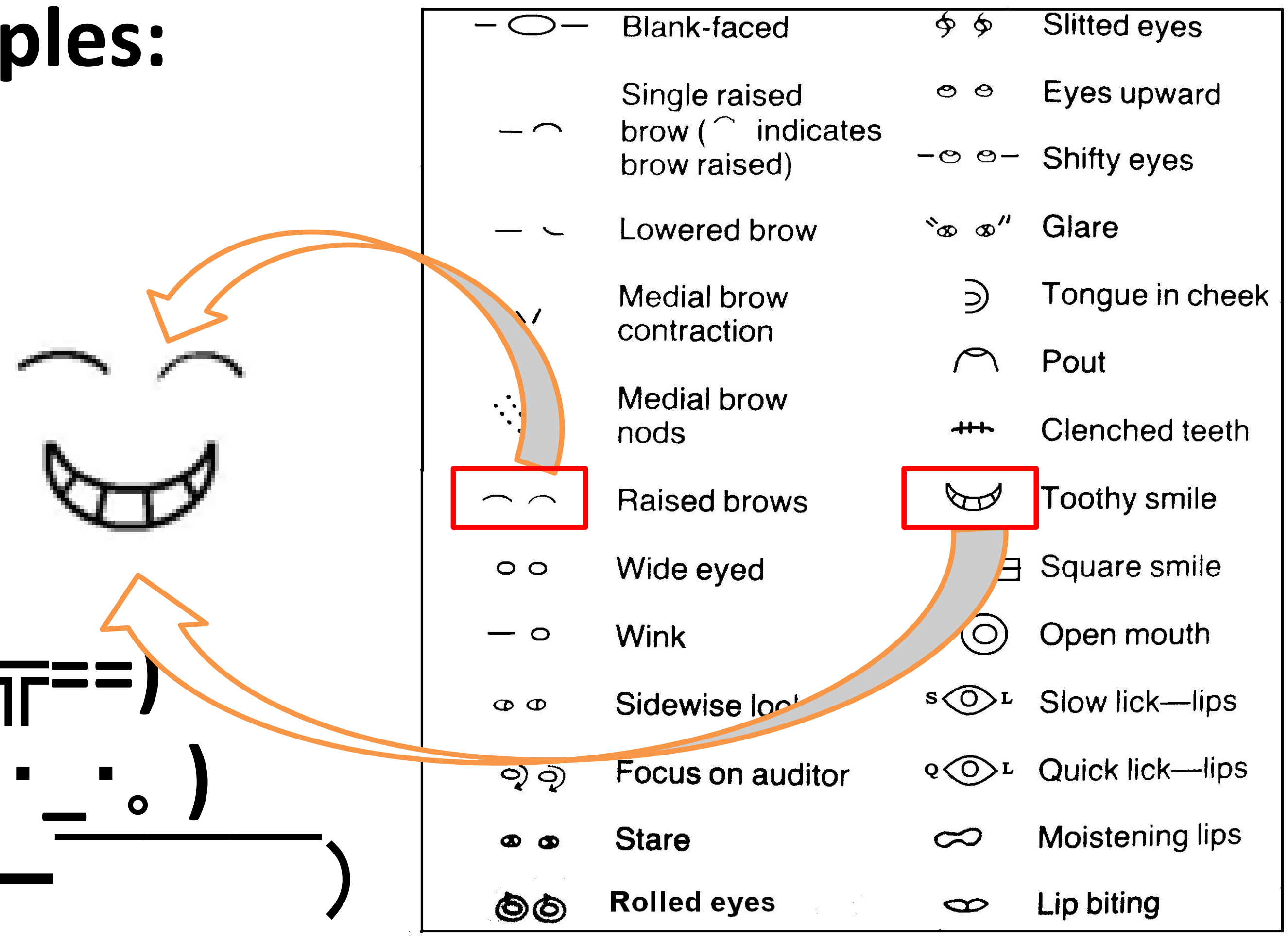
(°. °)

(^_-)y--~

(==╦==Д==╦==)

(｡･_･｡)人(｡･_･｡)

(*￣＿￣)



Emoticons - Definition

Some examples:

\(*^o^*)/

・゜・(/Д`;)・゜

(;^_^A

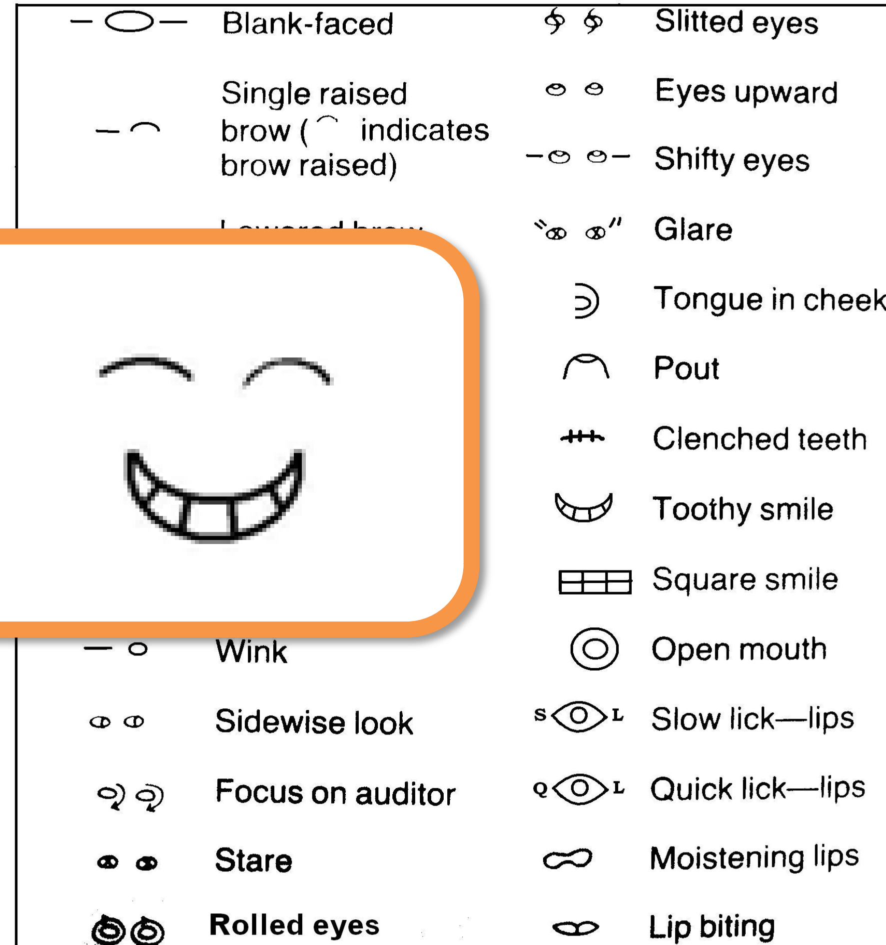
(°.　°)

(^_-)y-

(==╦==Д==╦==)

(｡･_･｡)人(｡･_･｡)

(*￣＿￣)



^_^ =

Emoticons - Definition

\(*^o^*)/

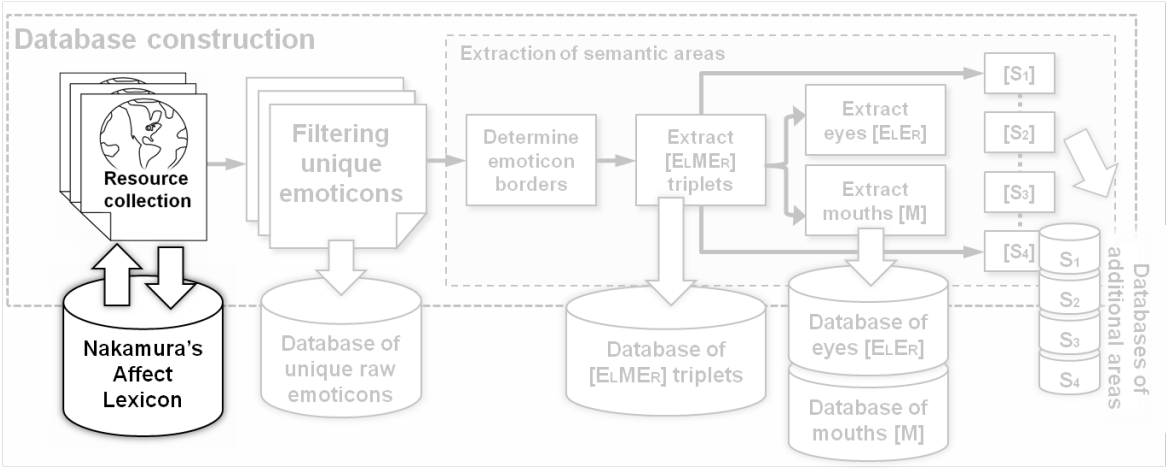
- Additional area: \
- Bracket: (
- Additional area: *
- Face: ^o^ (Eyes: ^ ^, Mouth: o)
- Additional area: *
- Bracket:)
- Additional area: /

Assumption:
Emoticons could be analyzed by dividing them to areas (kinemes)!

Database Construction

- Visited 7 online emoticon dictionaries:

1. *Face-mark Party,* 2. *Kaomo-jiya,*
3. *Kao-moji-toshokan,* 4. *Kaomoji-café,*
5. *Kaomoji Paradise,* 6. *Kaomojisyo and*
7. *Kaomoji Station.*



http://www.facemark.jp/facemark.htm,
http://kaomojiya.com/,
http://www.kaomoji.com/kao/text/,
http://kaomoji-cafe.jp/, http://rsmz.net/kaopara/,
http://matsucon.net/material/dic/,
http://kaosute.net/jisyo/kanjou.shtml

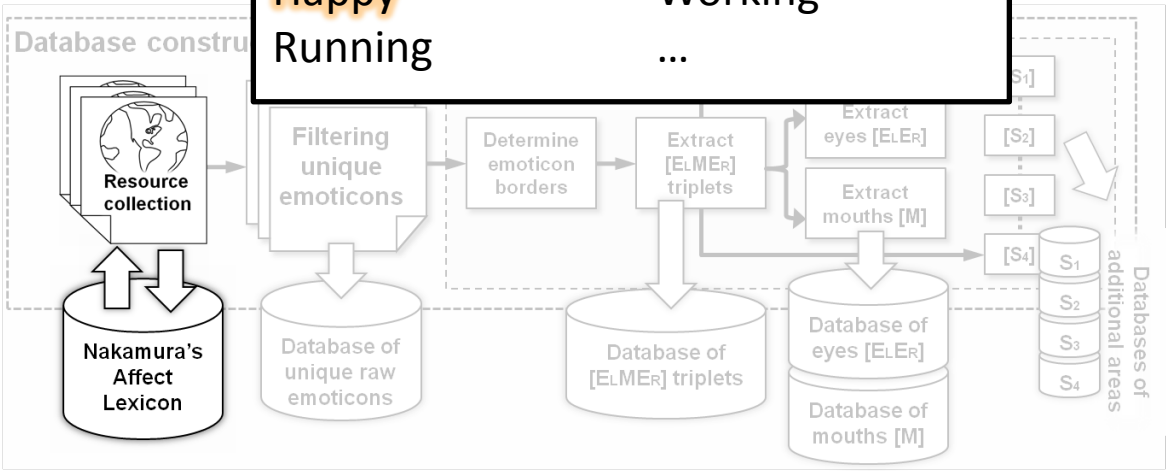
Database Construction

- Used an affect analysis system to select and categorize only emotion-related labels.

Emoticon Dictionary

Emoticon labels:

Hobbies	Screaming
Greetings	Joyful
Angry	Crying
Happy	Working
Running	…



Ptaszynski, M., Dybala, P., Rzepka, R. and Araki, K. 2009. Affecting Corpora: Experiments with Automatic Affect Annota-tion System - A Case Study of the 2channel Forum, In *Proceedings PACLING-09, pp. 223-228.*

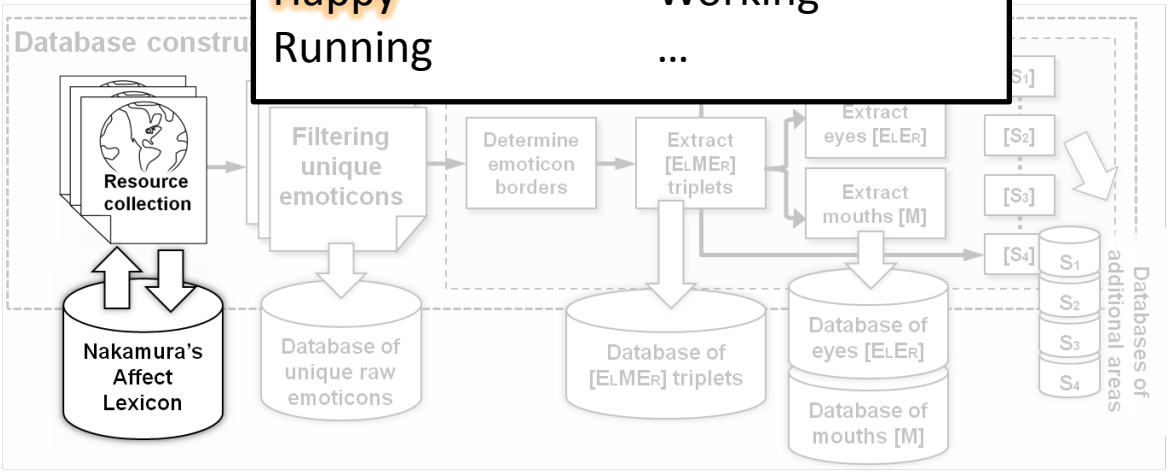
Database Construction

- Used an affect analysis system to select and categorize only emotion-related labels.
- Extract emoticons only from labels related to emotions

Emoticon Dictionary

Emoticon labels:

Hobbies	Screaming
Greetings	Joyful
Angry	Crying
Happy	Working
Running	…



\(*^o^*)/

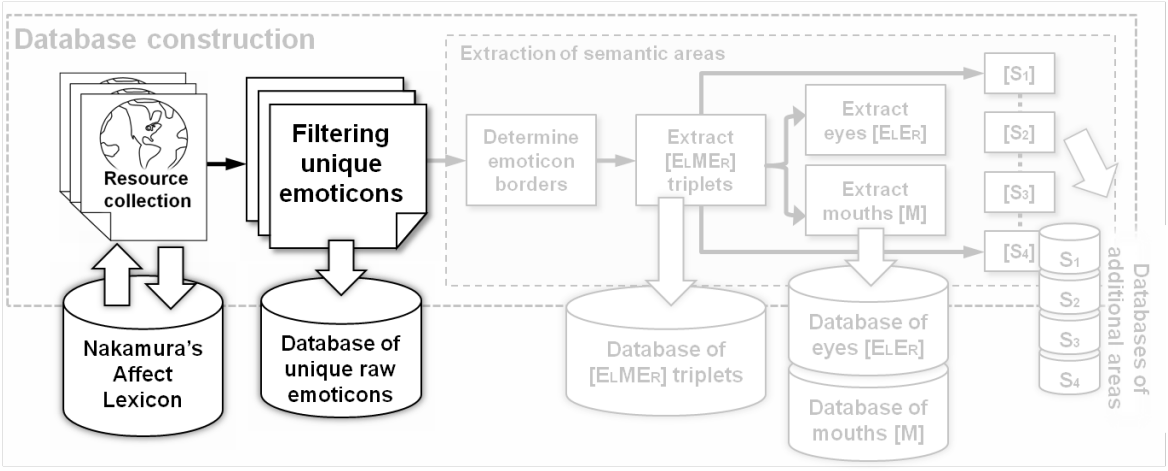
Ptaszynski, M., Dybala, P., Rzepka, R. and Araki, K. 2009. Affecting Corpora: Experiments with Automatic Affect Annota-tion System - A Case Study of the 2channel Forum, In *Proceedings PACLING-09, pp. 223-228.*

Database Construction

- Obtained 10,137 unique emoticons classified with emotion types.

joy, delight	liking, fondness	anger	surprise, amazement	sadness, gloom	excite-cite-ment	dis-like	shame, shyness	fear	relief	Over-all	Emoticons
3128	1988	1238	1227	1203	1124	704	526	179	99	**11416**	All extracted
1972	1972	1221	1196	1169	1120	698	511	179	99	**10137**	Unique
63%	99%	99%	97%	97%	99%	99%	97%	100%	100%	**89%**	Ratio

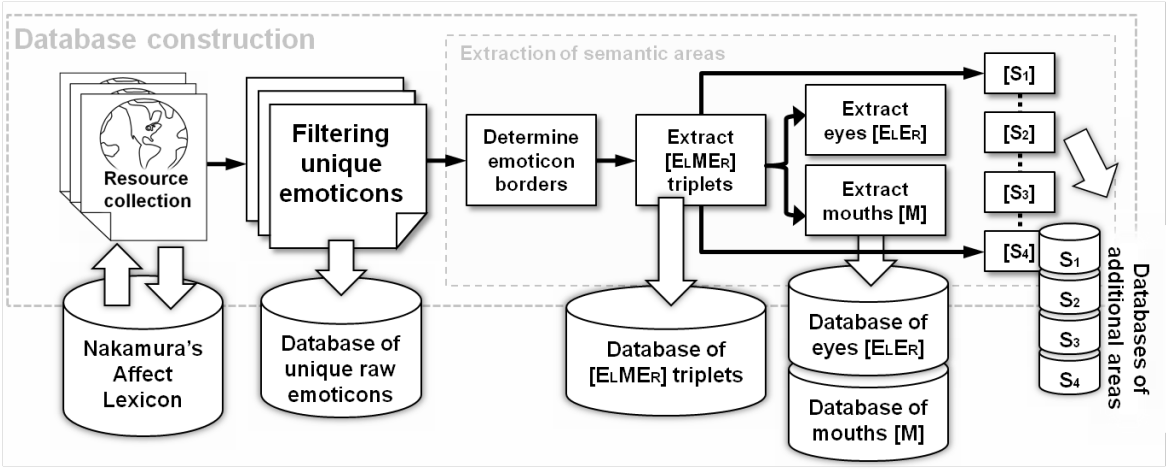
\(*^o^*)/



Database Construction

- Automatically divide emoticons into:
  - Eyes [E]: ^ ^
  - Mouths [M]: o
  - Additional areas (inside emoticon) [S]: * *
  - Additional areas (outside emoticon) [S]: \ /

\(*^o^*)/



Database Construction

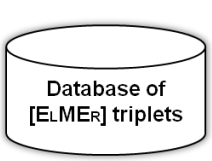
- We have a set of databases!
  - Raw emoticons
  - Triplets (E-M-E)
  - Eyes (E-E)
  - Mouths (M)
  - Additional (S)

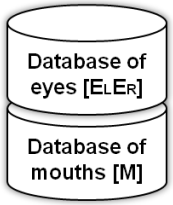
\(*^o^*)/

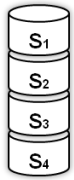
joy, delight	liking, fondness	anger	surprise, amazement	sadness, gloom	excite-cite-ment	dis-like	shame, shyness	fear	relief	Over-all	Emoticons
3128	1988	1238	1227	1203	1124	704	526	179	99	**11416**	All extracted
1972	1972	1221	1196	1169	1120	698	511	179	99	**10137**	Unique
63%	99%	99%	97%	97%	99%	99%	97%	100%	100%	**89%**	Ratio

areas	$E_L M E_R$	S_1	B_1	S_2	E_L E_R	M	S_3	B_2	S_4
joy, delight	1298	1469	--	653	349	336	671	--	2449
anger	741	525	--	321	188	239	330	--	1014
sadness,	702	350	--	303	291	170	358	--	730
fear	124	72	--	67	52	62	74	--	133
shame, shyness	315	169	--	121	110	85	123	--	343
liking, fondness	1079	1092	--	802	305	239	805	--	1633
dislike	527	337	--	209	161	179	201	--	562
excitement	670	700	--	268	243	164	324	--	1049
relief	81	50	--	11	38	26	27	--	64
surprise, amazement	648	405	--	231	183	154	279	--	860
overall	**6185**	**5169**	**--**	**2986**	**1920**	**1654**	**3192**	**--**	**8837**

Already annotated with emotion types!





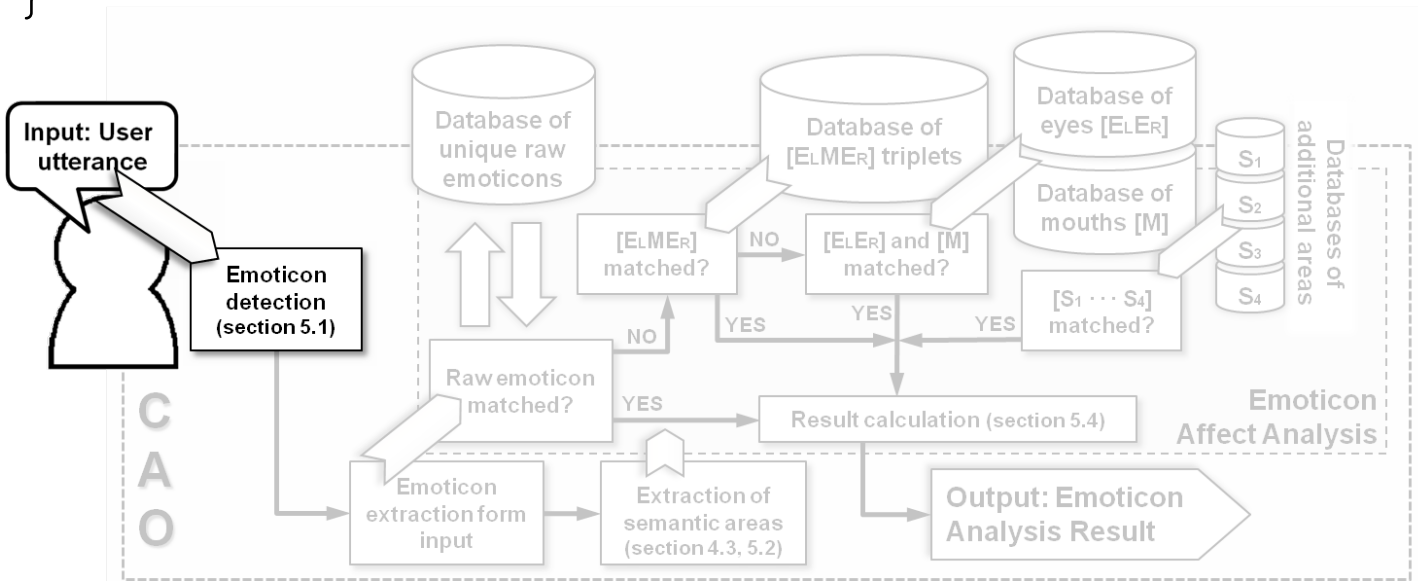


CAO – Emoticon Analysis System

- Constructed CAO system for emoticon analysis with these databases.

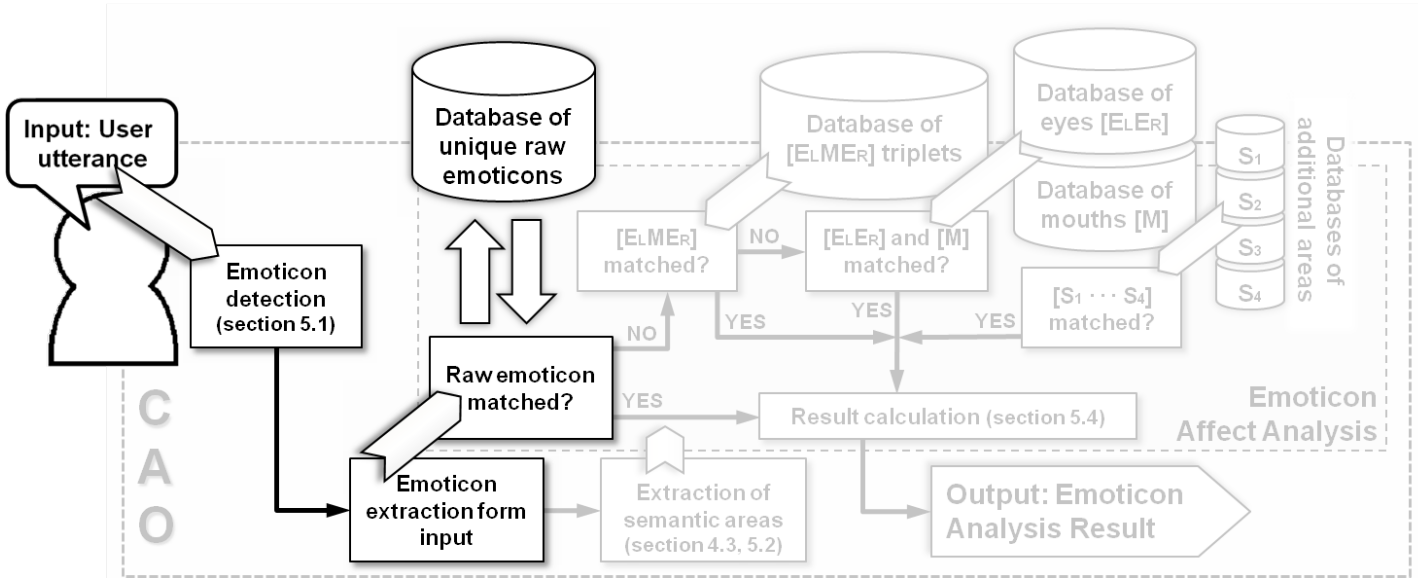
CAO – Emoticon Analysis System

- Emoticon detection in (any) input
  - Use 455 characters most frequently (>10 times) appearing in emoticons ($x_1, x_2, \ldots x_{455}$)
  - `If (any three x appear in a row) { there is an emoticon in input }`



CAO – Emoticon Analysis System

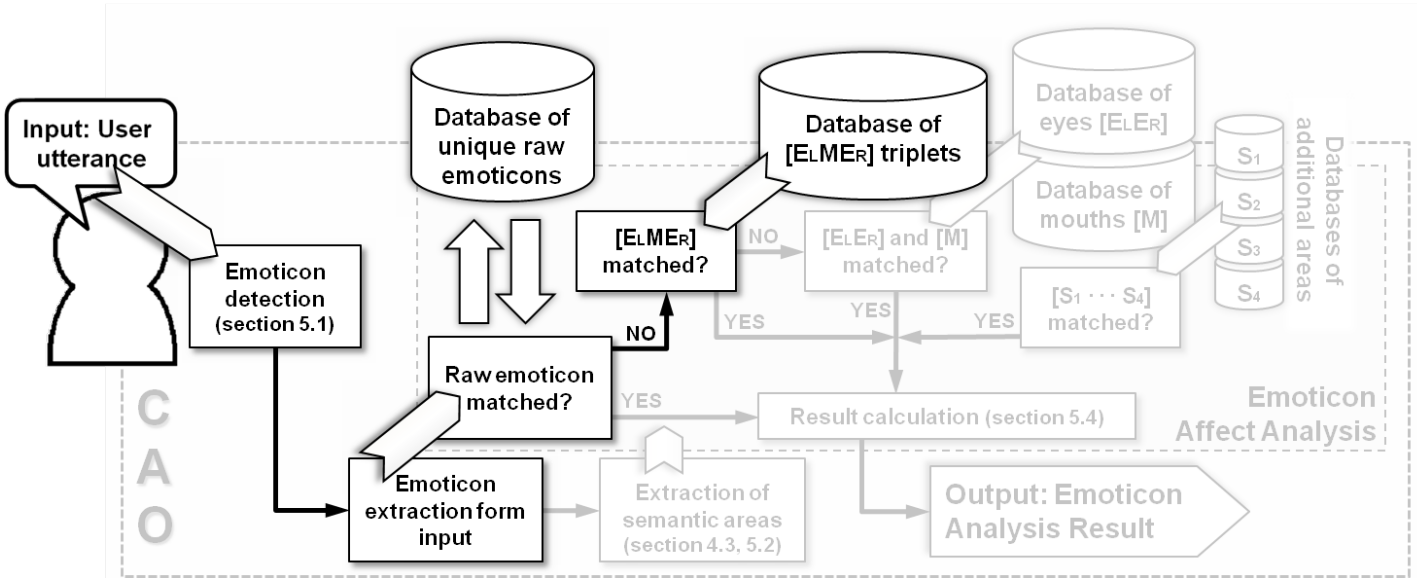
- Emoticon extraction from input (+ affect analysis)
  - Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)



CAO – Emoticon Analysis System

- Emoticon extraction from input (+ affect analysis)
  - Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)
    2. Looking for a triplet (+checking emotion labels)

If no “raw” emoticon

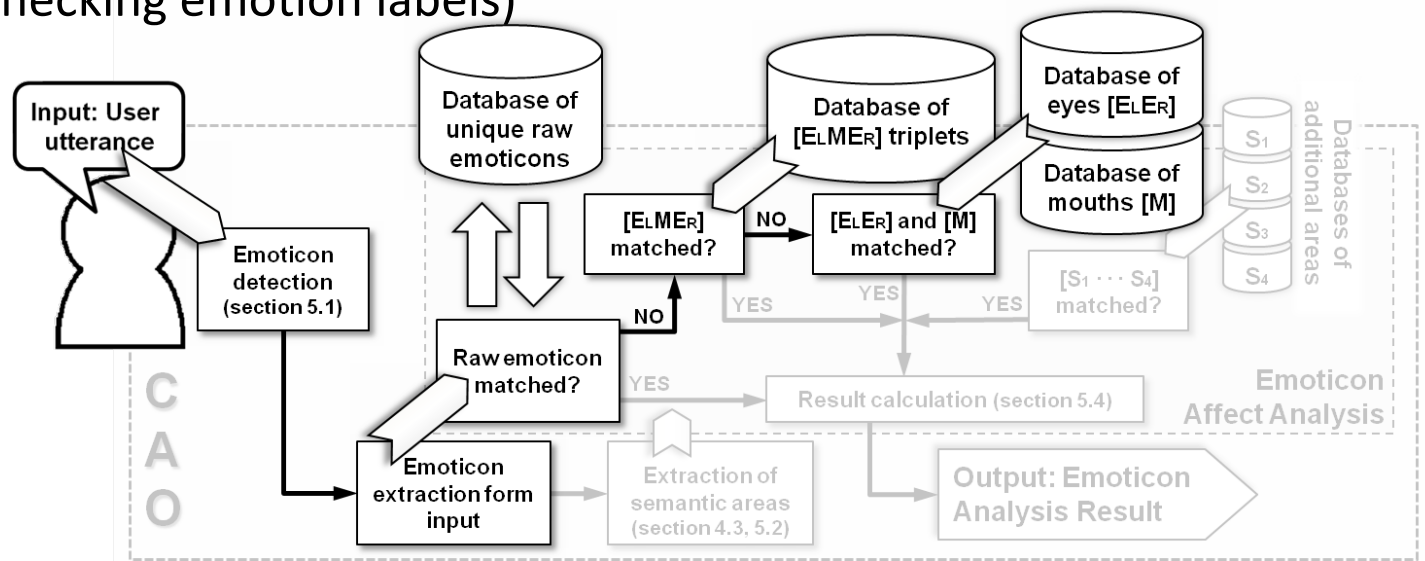


CAO – Emoticon Analysis System

- Emoticon extraction from input (+ affect analysis)
  - Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)
    2. Looking for a triplet (+checking emotion labels)
    3. Checking all combinations of triplets (eyes x mouth*) (+checking emotion labels)

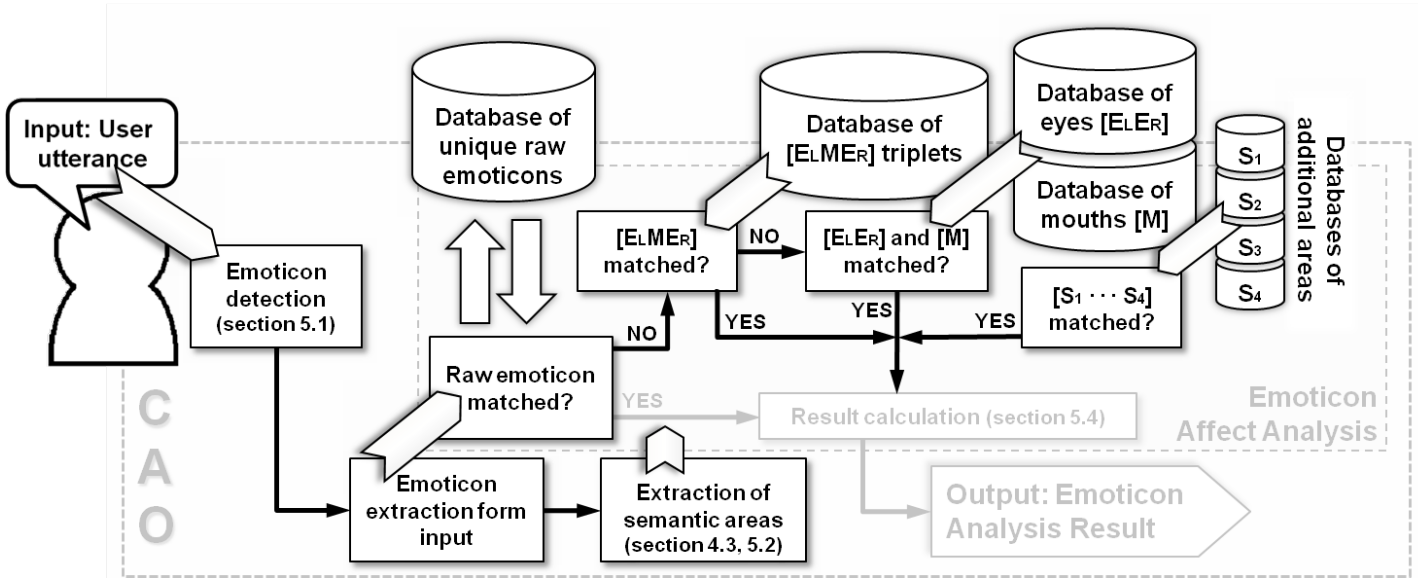
If no triplet

*)Eyes=1,920
Mouths=1,654
All combinations:
ExM=3,175,680



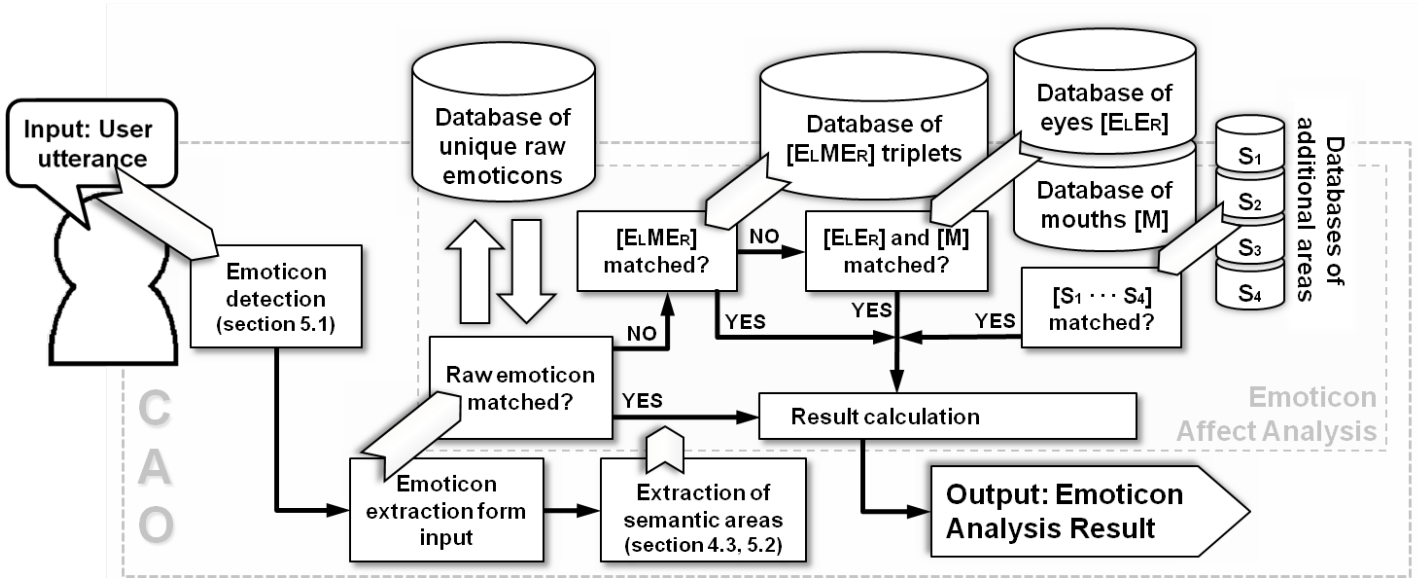
CAO – Emoticon Analysis System

- Emoticon extraction from input
  - Finally:
    - Extract additional areas (+checking emotion labels)



CAO – Emoticon Analysis System

- Emoticon extraction from input
  - Finally:
    - Extract additional areas (+checking emotion labels)
  - Summarize scores (to determine emotion types statistically most probable for this emoticon)



Evaluation of CAO

- Test set
  - A large corpus of blogs from: Ameba Blog*
  - 354,288,529 Japanese sentences in
  - 12,938,606 downloaded and parsed web pages
  - written by 60,658 unique bloggers

*) www.ameblo.co.jp

Evaluation of CAO

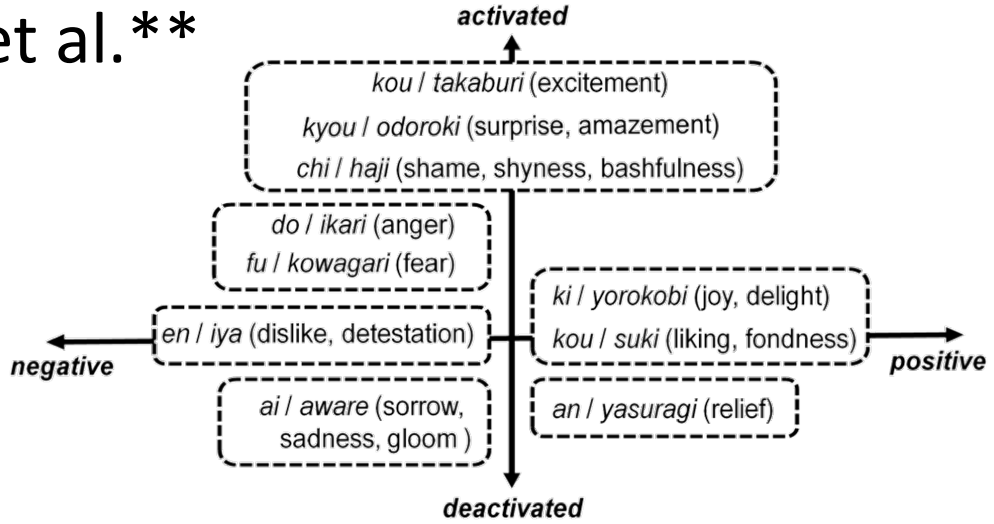
- Randomly extracted 1000 middle-sized* sentences as the test set
  - 418 of those sentences included emoticons.
    - annotate the sentences with 42 people (10 sentences per 1 person)
      Question: What emotion was expressed in the sentence?
    - annotate emoticons from the sentences (different samples than in sentences)
      Question: What emotion could be expressed with this emoticon?
    - Answers (emotion type, random order): *a) System's; b) Similar**; c) Completely different; d) Other (from the seven remaining);*

Test Set Gold standard

*)20-50 characters in Japanese **) From the same affect space in two-dimensional model of affect

Evaluation of CAO

- Estimation of:
  - Emotion types (10 types)
  - General emotive features (valence and activation)* adjusted to Japanese like in Ptaszynski et al.**



*) Russell, J. A. 1980. A circumplex model of affect, *J. of Personality and Social Psychology, 39(6), pp. 1161-1178.*

**) Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. and Araki, K. 2009. Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States, In *Proceedings of IJCAI-09, pp. 1469-1474.*

Evaluation of CAO

- Results

Detection			
		System	
		Emoticon	No emoticon
Users	Emoticon	394	24
	No emoticon	0	582
No. of agreements=976 (97.6%), Kappa=0.95			

In 24/418 cases there were no 3 usual chars in a row

Evaluation of CAO

- Results

Extraction		
R	P	F-score
94.3%	100%	97.1%
$(\frac{394}{418})$	$(\frac{394}{394})$	$2\frac{P*R}{P+R}$

Errors only for the undetected emoticons

Evaluation of CAO

- Results
- Emotion Estimation on Separate Emoticons
  - Emotion types: 93.54%
  - General emotive features: 97.39%

Accuracy in determining probable emotion types a certain emoticon could be used to express

Evaluation of CAO

- Results
- Emotion Estimation on Sentences
  - Emotion types: 80.2%
  - General emotive features: 94.63%

Accuracy in determining emotion types expressed in a sentence, only with the use of emoticon*

*) a sentence needs to contain at least one emoticon

Evaluation of CAO

- Results
- Emotion Estimation on Sentences
  - Emotion types: 80.2%
  - General emotive features: 94.63%

1. The results were worse because meaning in sentences is conveyed also through lexical channel; but,
2. Results for general features were high → People sometimes misinterpret specific emotion type, but rarely valence/activation;

Accuracy in determining emotion types expressed in a sentence, only with the use of emoticon*

*) a sentence needs to contain at least one emoticon

Conclusions

- Presented a prototype system for automatic affect analysis of Eastern type emoticons, CAO.
- Inspired by Theory of Kinesics
- Gathered database of +10,000 emoticons and (almost) automatically expanded it to +3 mln.

Conclusions

- CAO is capable of:
  - Detecting emoticons in any input
  - Extracting emoticons form input
  - Dividing emoticons into semantic areas (eyes, mouths, etc.)
  - Estimating potential emotion types expressed by emoticons.
  - Affect analysis of sentences including emoticons
- CAO got almost ideal results in all tasks.

Future Work

Possible applications:

- Affect analysis/annotation of corpora
- Emotion detecting in online communication
  - Support for Internet messengers, blog services, forums, etc.
- Sentiment analysis (when looking only at valence)
- Detecting irony*

*) Carvalho, P., Sarmento, L., Silva, M. J., and de Oliveira, E. 2009. Clues for detecting irony in user-generated contents: oh..!! it's "so easy" ;-). In Proceeding of the 1st international CIKM Workshop on Topic-Sentiment Analysis For Mass Opinion (Hong Kong, China, November 06 - 06, 2009)

Thank you for your attention!

Read more in: **"A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics"**

http://www.computer.org/portal/web/tac

Details on Extraction of Emoticon Areas

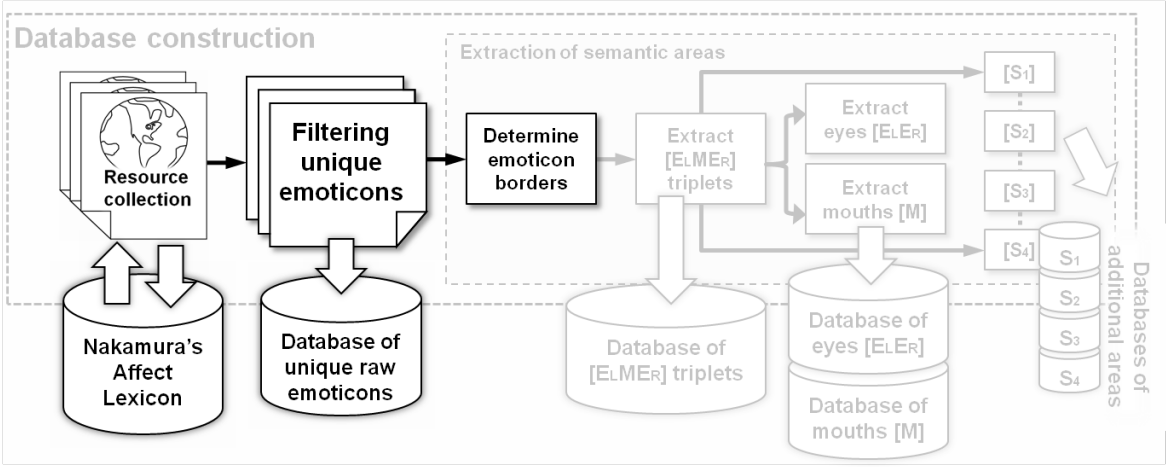
In: Database Construction

Database Construction

- Determined all possible emoticon borders:

```
2bit chars: (, 【, 〔, 〈; 1-bit chars: (, [,
  <, |; [none]
2bit chars: ), 】, 〕, 〉; 1-bit chars: ), ],
  >, |; [none]
```

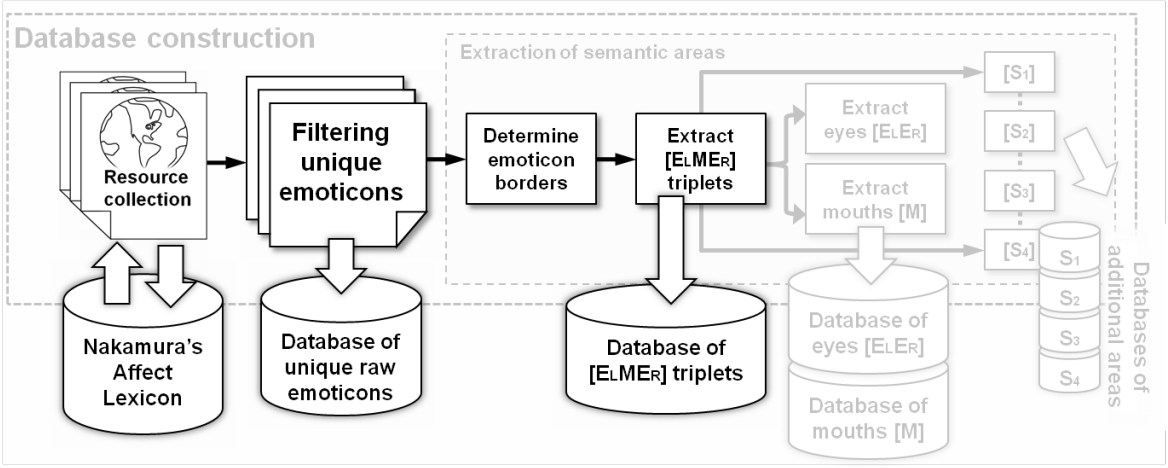
\(*^o^*)/



Database Construction

- Extract eye-mouth-eye triplets
  - `Get rid of what is behind brackets (inclusively with brackets)`
  - Get rid of additional areas from within emoticons (the only detail done manually)
- Make a database of emoticon triplets

\(*^o^*)/

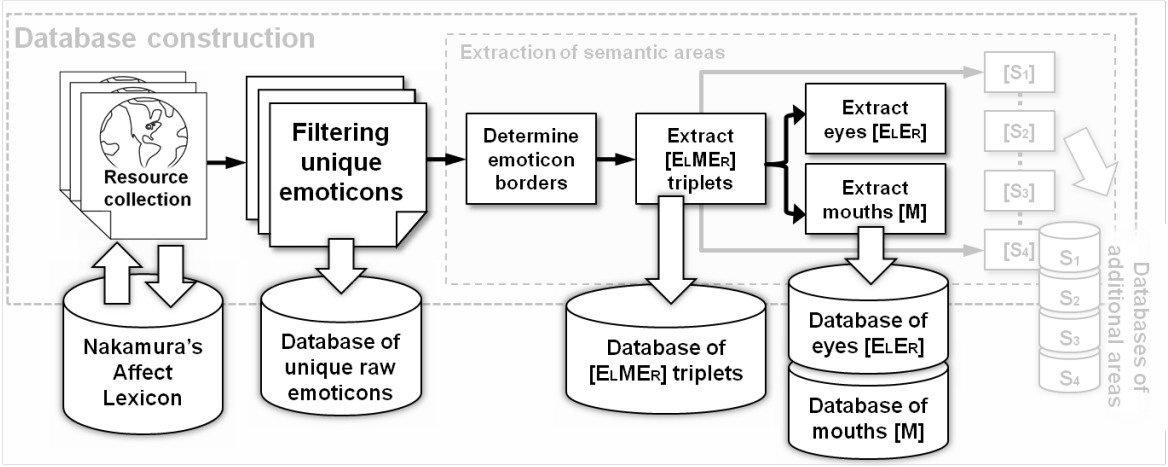


Database Construction

- Extract eyes and mouths
  - If an eye has more than 1 character, both eyes are the same;

```
if (n characters from left and right
match) {n=eye};
ifelse (take n-1,n-2,n-3,…)
```

\(*^o^*)/

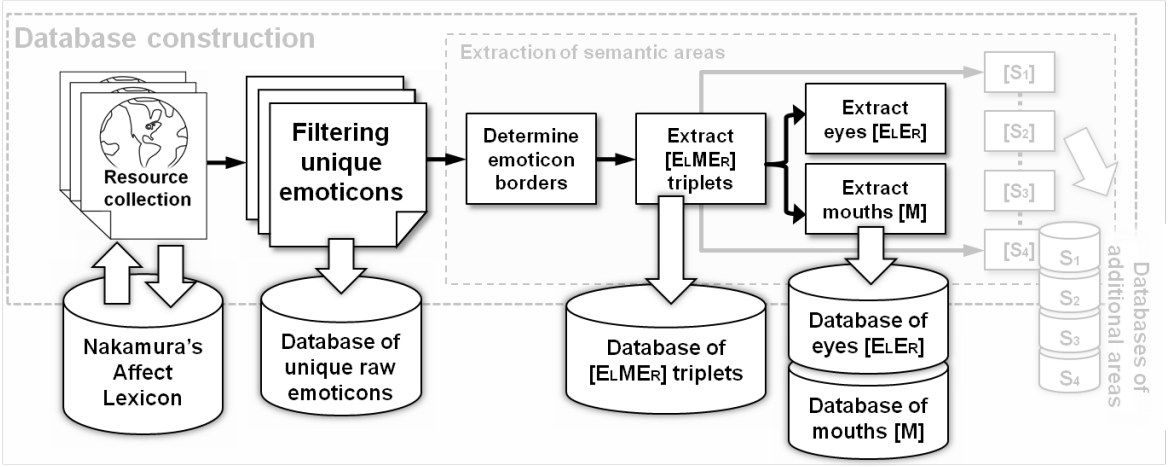


Database Construction

- Extract eyes and mouths
  - If an eye has 1 character, eyes could be the same or different;

```
else(
        take 1 char. from left and right as eyes;
        mouth is what is left inside;
        )
```

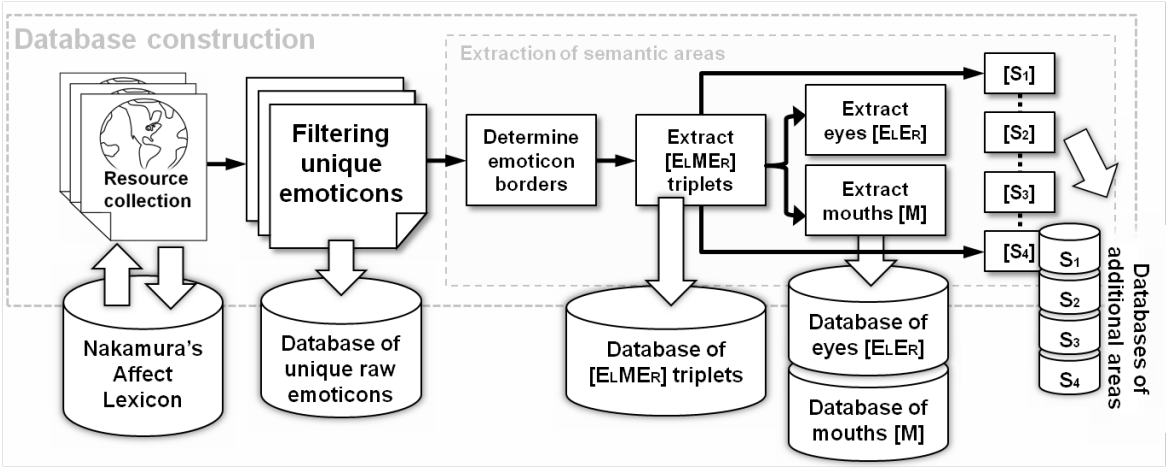
\(*^o^*)/



Database Construction

- Extract additional areas
  - Localize and extract additional areas
  - Make database of additional areas

\(*^o^*)/



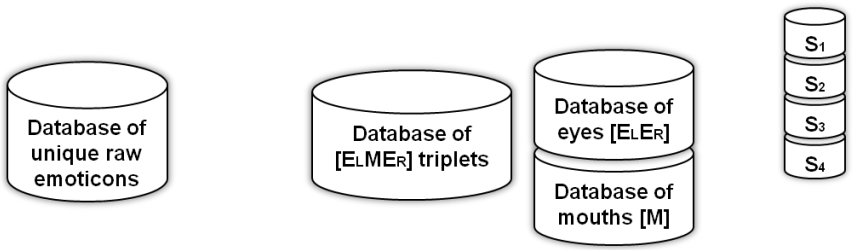
Database Construction

- We have a database!

\(*^o^*)/

joy, delight	liking, fondness	anger	surprise, amazement	sadness, gloom	excite-cite-ment	dis-like	shame, shyness	fear	relief	Over-all	Emoticons
3128	1988	1238	1227	1203	1124	704	526	179	99	**11416**	All extracted
1972	1972	1221	1196	1169	1120	698	511	179	99	**10137**	Unique
63%	99%	99%	97%	97%	99%	99%	97%	100%	100%	**89%**	Ratio

areas	$E_L M E_R$	S_1	B_1	S_2	$E_L E_R$	M	S_3	B_2	S_4
joy, delight	1298	1469	--	653	349	336	671	--	2449
anger	741	525	--	321	188	239	330	--	1014
sadness,	702	350	--	303	291	170	358	--	730
fear	124	72	--	67	52	62	74	--	133
shame, shyness	315	169	--	121	110	85	123	--	343
liking, fondness	1079	1092	--	802	305	239	805	--	1633
dislike	527	337	--	209	161	179	201	--	562
excitement	670	700	--	268	243	164	324	--	1049
relief	81	50	--	11	38	26	27	--	64
surprise, amazement	648	405	--	231	183	154	279	--	860
overall	**6185**	**5169**	**--**	**2986**	**1920**	**1654**	**3192**	**--**	**8837**

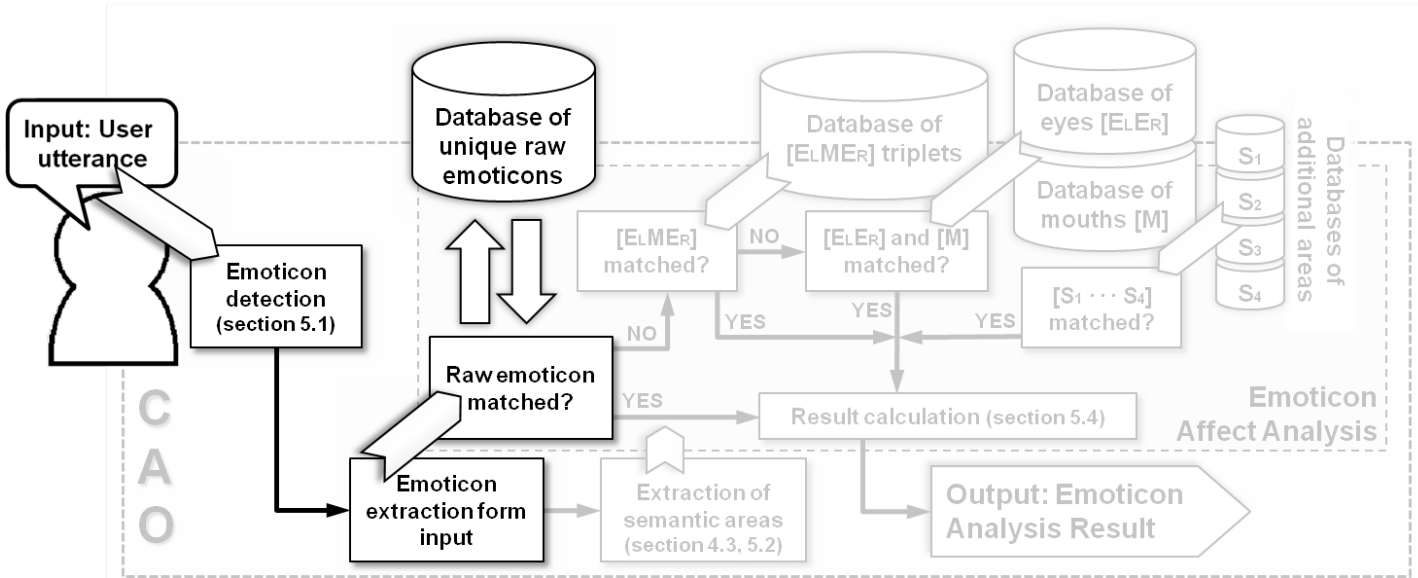


Details on Affect Analysis of Emoticons

In: CAO – Emoticon Analysis System

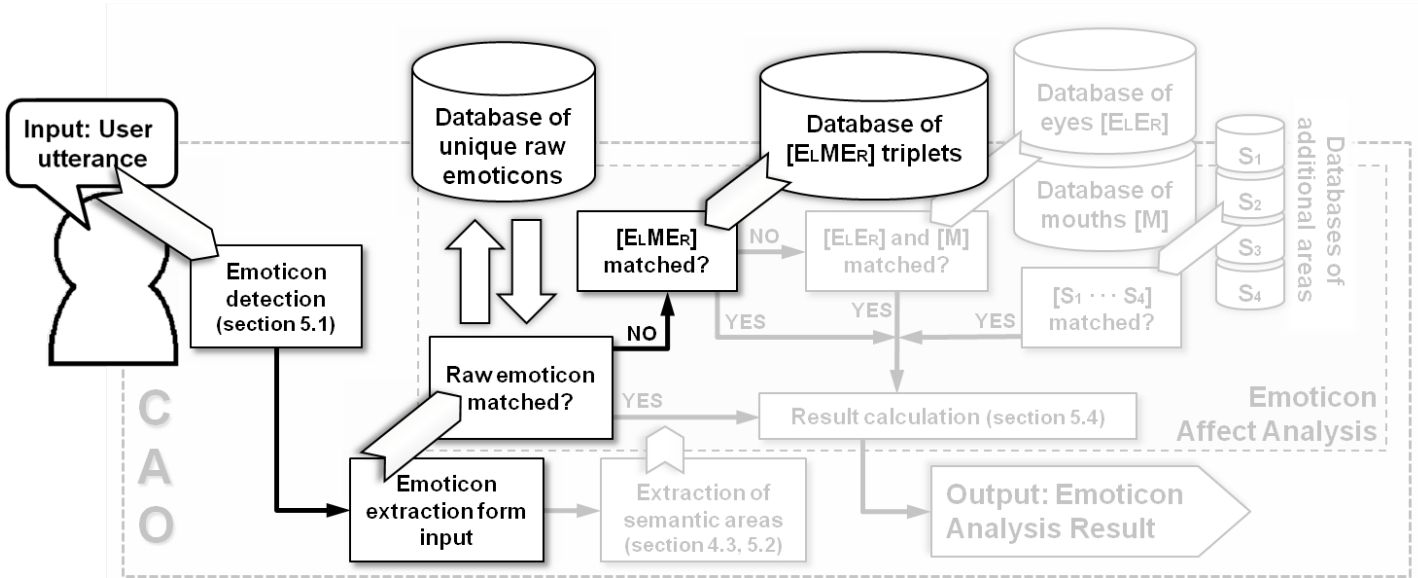
CAO – Emoticon Analysis System

- Emoticon affect analysis (along with extraction)
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons



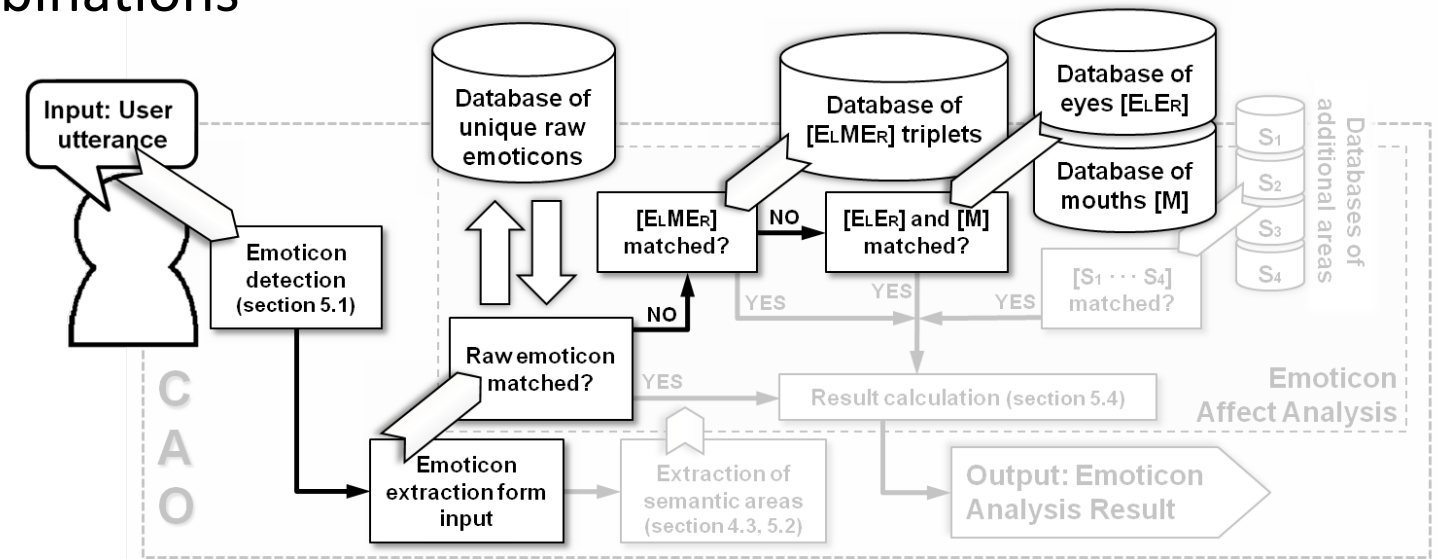
CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons
    - For [2.]: Check emotion types annotated on triplets



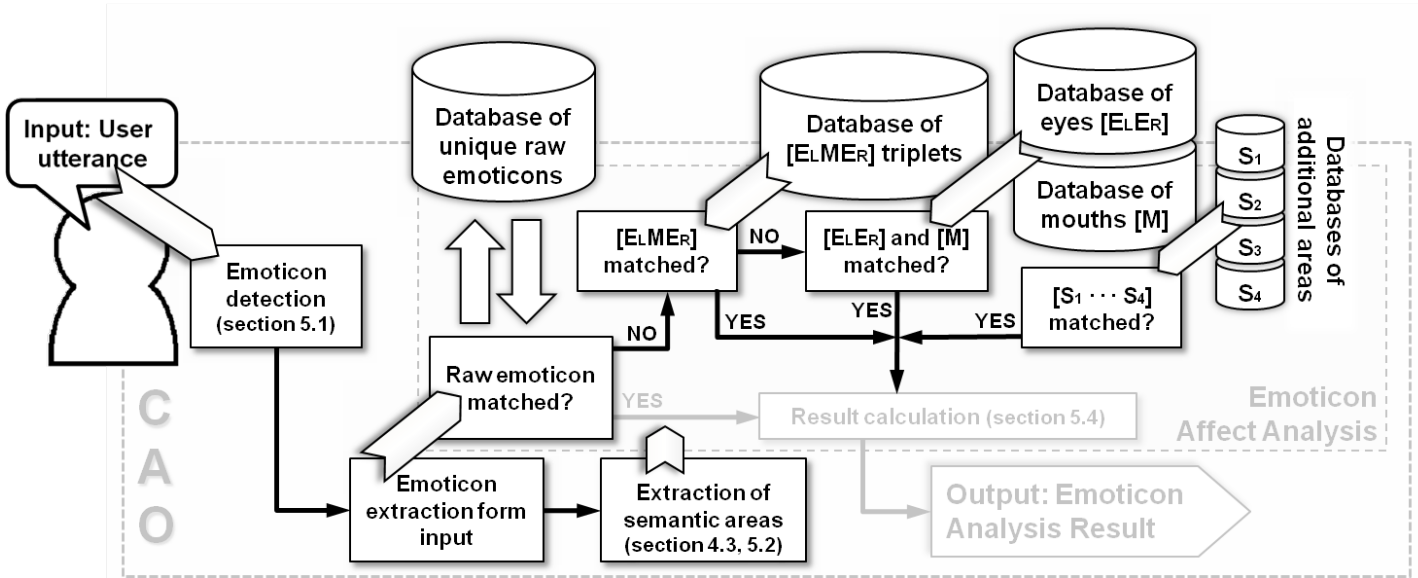
CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons
    - For [2.]: Check emotion types annotated on triplets
    - For [3.]: Check emotion types annotated on separate ExM combinations



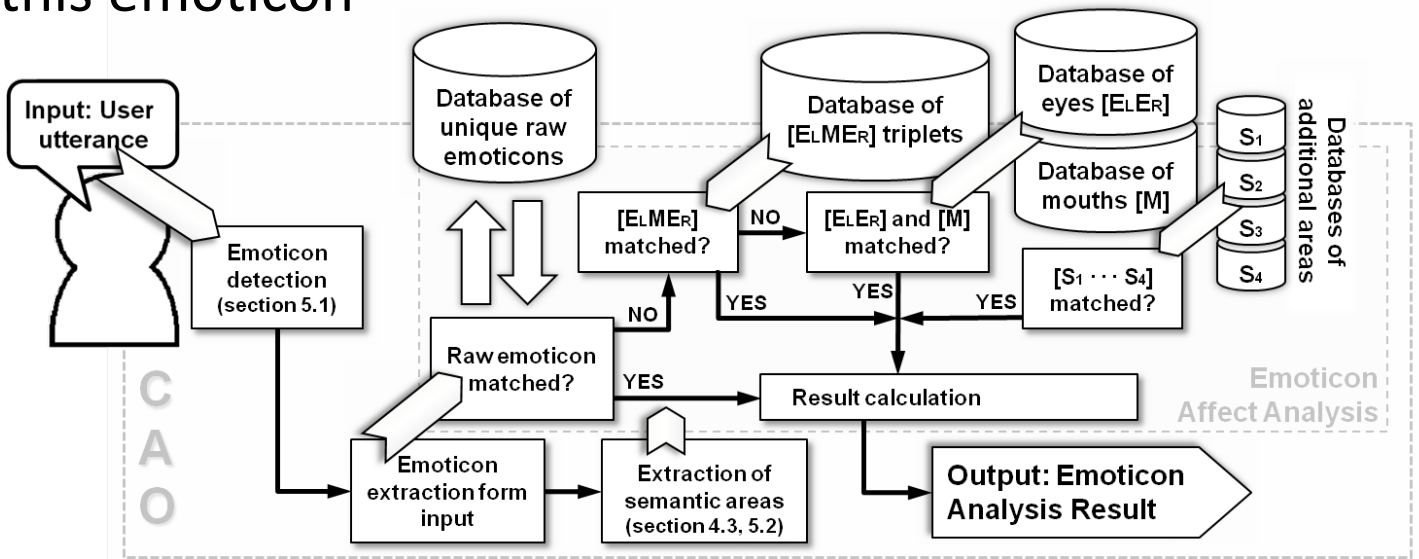
CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Finally
    - Check emotion types annotated on additional areas



CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Finally
    - Check emotion types annotated on additional areas
    - Summarize score
    - Give output: list of emotions most probably expressed with this emoticon



Summarizing scores

In: Evaluation of CAO

Evaluation of CAO

- Summarizing scores
  - Occurrence
    - Sum of all emotion types found for all elements
  - Frequency
    - Sum for each element divided by number of all elements in each database
  - Unique frequency
    - Sum for each element divided by number of unique elements in each database

Evaluation of CAO

Detailed Description

Evaluation of CAO

- Training Set*
  - Raw emoticon database (Tr.S. gold standard)
    - Take emoticon from a database (e.g. from “joy”)
    - Process
    - Check result with gold standard

joy, delight	liking, fondness	anger	surprise, amazement	sadness, gloom	excite-cite-ment	dis-like	shame, shyness	fear	relief	Over-all	Emoticons
3128	1988	1238	1227	1203	1124	704	526	179	99	**11416**	All extracted
1972	1972	1221	1196	1169	1120	698	511	179	99	**10137**	Unique
63%	99%	99%	97%	97%	99%	99%	97%	100%	100%	**89%**	Ratio

*) In training set evaluation we matched only triplets and all possible; matching also raw would give all 100%

Evaluation of CAO

- Training Set*
  - Raw emoticon database (Tr.S. gold
    - Take emoticon from a
    - Process
    - Check result with gold standard
    - Ranking:
      1. **Occurrence**
      2. **Unique Frequency**
      3. **Frequency**

      (differences not significant=all equally good)

Emotion type	CAO: Occur-rence	Freq-uency	Unique Freq-uency
anger	**0.811**	0.771	0.767
dislike	0.631	**0.800**	0.719
excitement	0.786	0.769	**0.797**
fear	0.451	**0.936**	0.858
fondness	**0.915**	0.778	0.783
joy	**0.944**	0.802	0.860
relief	0.600	**0.990**	0.985
shame	0.706	**0.922**	0.910
sorrow	**0.814**	0.809	0.791
surprise	0.862	0.866	**0.874**
All approx.	**0.852**	0.804	0.818

1 3 2

Evaluation of CAO

- Test set
  - A large corpus of blogs from: Ameba Blog*
  - 354,288,529 Japanese sentences in
  - 12,938,606 downloaded and parsed web pages
  - written by 60,658 unique bloggers
- Randomly extracted 1000 middle-sized** sentences as the test set
  - 418 of those sentences included emoticons.

Test Set Gold standard

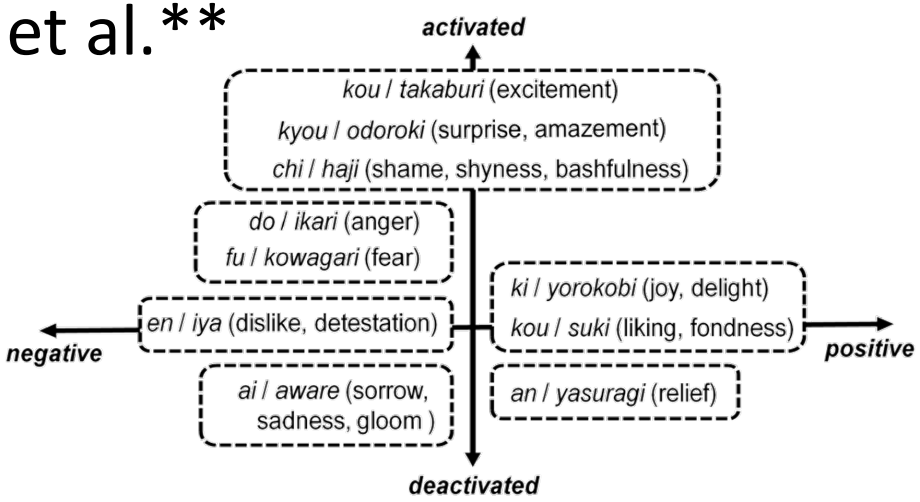
- annotate the sentences with 42 people (10 sentences per 1 person)
  Question: What emotion was expressed in the sentence?
- annotate emoticons from the sentences (different samples than in sentences)
  Question: What emotion could be expressed with this emoticon?

*) www.ameblo.co.jp

**)20-50 characters in Japanese

Evaluation of CAO

- Estimation of:
  - Emotion types (specific)
  - General emotive features (valence and activation)* adjusted to Japanese like in Ptaszynski et al.**



*) Russell, J. A. 1980. A circumplex model of affect, *J. of Personality and Social Psychology, 39(6), pp. 1161-1178.*

**) Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. and Araki, K. 2009. Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States, In *Proceedings of IJCAI-09, pp. 1469-1474.*

Evaluation of CAO

- Results

Detection			
		System	
		Emoticon	No emoticon
Users	Emoticon	394	24
	No emoticon	0	582
No. of agreements=976 (97.6%), Kappa=0.95			

Evaluation of CAO

- Results

Extraction		
R	P	F-score
94.3%	100%	97.1%
$(\frac{394}{418})$	$(\frac{394}{394})$	$2\frac{P*R}{P+R}$

Evaluation of CAO

- Results

Emotion Estimation on Separate Emoticons

CAO					
Occurrence (3)		Frequency (2)		Unique Frequency (1)	
Types	2D space	Types	2D space	Types	2D space
0.891472	0.966778	0.934319	0.971044	**0.935364**	**0.973925**

Emotion Estimation on Sentences

CAO					
Occurrence		Frequency		Unique Frequency	
Types	2D space	Types	2D space	Types	2D space
0.755171	0.908911	0.800896	0.940582	**0.802012**	**0.946291**

1. **Unique Frequency**
2. **Frequency**
3. **Occurrence**

Comparing CAO to other systems

In: Evaluation of CAO

Evaluation of CAO

- Comparing CAO to other systems

Research → (approach) Capability ↓	Tanaka et al. (2005) (kernel methods)	Yamada et al. (2007) (n-grams)	Kawakami (2008) (database)	CAO (theory of kinesics)
1. Detection whether input equals emoticon	X	X	X	O
2. Detection of emoticon in sentence input	O (included in 3.)	X	X	O
3. Extraction of emoticon from any string of characters	O	X	X	O
4. Division into semantic areas	X	X	X	O
5. Database coverage	1,075	693	31	10,137 (expanded automatically to over 3 million)
6. Classification of emotion types	6 types (BBS-based;	7 types Subjective (Subjective)	6 types (Subjective)	10 types (Language/ Culture Based)
7. Emotion estimation of separate emoticons	O (included in 8.)	O	O	O
8. Affect Analysis of sentences with emoticons	O	X	X	O

Evaluation of CAO

- Comparing CAO to other systems
  - In Training set: Comparison with Yamada et al. (2007)

Upgraded with our database and emotion classification

T. Yamada, S. Tsuchiya, S. Kuroiwa, F. Ren, "Classification of Facemarks Using N-gram", International Conference on NLP and Knowledge Engineering, pp. 322-327, 2007.

Evaluation of CAO

- Comparing CAO to other systems
  - In Training set: Comparison with Yamada et al. (2007) (their best < our worst)

Emotion type	Yamada et al (2007) 1-gram	2-gram	3-gram	CAO: Occur-rence	Freq-uency	Unique Freq-uency
anger	0.702	0.815	**0.877**	**0.811**	0.771	0.767
dislike	0.661	0.809	**0.919**	0.631	**0.800**	0.719
excitement	0.700	0.789	**0.846**	0.786	0.769	**0.797**
fear	**0.564**	0.409	0.397	0.451	**0.936**	0.858
fondness	**0.452**	0.436	0.448	**0.915**	0.778	0.783
joy	0.623	0.792	**0.873**	**0.944**	0.802	0.860
relief	**1.000**	0.999	**1.000**	0.600	**0.990**	0.985
shame	0.921	0.949	**0.976**	0.706	**0.922**	0.910
sorrow	0.720	0.861	**0.920**	**0.814**	0.809	0.791
surprise	0.805	0.904	**0.940**	0.862	0.866	**0.874**
All approx.	0.675	0.751	**0.802**	**0.852**	0.804	0.818

T. Yamada, S. Tsuchiya, S. Kuroiwa, F. Ren, "Classification of Facemarks Using N-gram", International Conference on NLP and Knowledge Engineering, pp. 322-327, 2007.

Evaluation of CAO

- Comparing CAO to other systems
  - In Test set: their best < our worst (or 2nd worst)

Emotion Estimation on Separate Emoticons

Yamada et al. (2007)			CAO					
			Occurrence		Frequency		Unique Frequency	
1-gram	2-gram	3-gram	Types	2D space	Types	2D space	Types	2D space
0.721347	0.865117	**0.877049**	0.891472	0.966778	0.934319	0.971044	**0.935364**	**0.973925**

Emotion Estimation on Sentences

Yamada et al. (2007)			CAO					
			Occurrence		Frequency		Unique Frequency	
1-gram	2-gram	3-gram	Types	2D space	Types	2D space	Types	2D space
0.685714	**0.797659**	0.714819	0.755171	0.908911	0.800896	0.940582	**0.802012**	**0.946291**

T. Yamada, S. Tsuchiya, S. Kuroiwa, F. Ren, "Classification of Facemarks Using N-gram", International Conference on NLP and Knowledge Engineering, pp. 322-327, 2007.

Statistical significance of results

Statistical significance or differences in training set evaluation

Emotion type	Yamada et al (2007) improved			CAO:				
	1-gram	2-gram	3-gram	Occurrence	Frequency	Unique Frequency	Position	Unique Position
anger	0.702	0.815	**0.877**	**0.811**	0.771	0.767	0.476	0.476
dislike	0.661	0.809	**0.919**	0.631	**0.8**	0.719	0.556	0.591
excitement	0.7	0.789	**0.846**	0.786	0.769	**0.797**	0.56	0.516
fear	**0.564**	0.409	0.397	0.451	**0.936**	0.858	0.652	0.671
fondness	**0.452**	0.436	0.448	**0.915**	0.778	0.783	0.46	0.389
joy	0.623	0.792	**0.873**	**0.944**	0.802	0.86	0.522	0.421
relief	**1**	0.999	**1**	0.6	**0.99**	0.985	0.599	0.621
shame	0.921	0.949	**0.976**	0.706	**0.922**	0.91	0.538	0.566
sorrow	0.72	0.861	**0.92**	**0.814**	0.809	0.791	0.553	0.52
surprise	0.805	0.904	**0.94**	0.862	0.866	**0.874**	0.52	0.523
All approx.	**0.675**	**0.751**	**0.802**	**0.852**	**0.804**	**0.818**	**0.517**	**0.469**

X	Y	Z	A	B	C

XvsY	not sign	AvsB	not sign
XvsZ	not sign	AvsC	not sign
YvsZ	not sign	BvsC	not sign

In most cases differences were not significant (results equally good).

	A	B	C
X	N	Y (<5%)	Y (<5%)
Y	N	N	N
Z	N	N	N

This means that although in training set Occurrence scored higher, Freq and UniqFreq were more probable to achieve better results in test set evaluation.