

How to find love in the Internet?

Applying Web mining to
affect recognition from textual input



Michal PTASZYNSKI
ptaszynski@media.eng.hokudai.ac.jp

Pawel DYBALA Wenhan SHI Rafal RZEPKA Kenji ARAKI

Language Media Laboratory
Graduate School of Information Science and Technology
Hokkaido University



北海道大学
HOKKAIDO UNIVERSITY



荒木研究室
ARAKI LABORATORY

1. Introduction
2. Affect analysis system of textual input in Japanese
3. Web mining technique for emotion types extraction
4. Two Perspective Evaluation
 1. “Cartesian”
 2. “Freudian”
5. Results
6. Conclusions and Future work

Introduction

Empirical methods for NLP:

- information extraction
- language and dialogue modeling
- discourse and dialogue processing
- tagging
- word sense disambiguation

Raison d'etre of Empirical methods for NLP:

- research in linguistics
- broadening our understanding of language
- using this knowledge in practice
- implementing language understanding in the machines (computing intelligence)



Computing Intelligence – Computing what?

...Intelligence?

“Ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought.”

American Psychological Association, 1995

Perloff, R.; Sternberg, R.J.; Urbina, S. (1996). "Intelligence: knowns and unknowns". American Psychologist

“Ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. (...) Capability for comprehending our surroundings—*"catching on", "making sense" of things, or "figuring out" what to do.*”

Mainstream Science on Intelligence, 1994

Gottfredson, L.S. (1997). "Foreword to "intelligence and social policy"". Intelligence 24 (1): 1–12



Intelligence – only one, or one of many?

Howard Gardner – “IQ tells you nothing!”. (Theory of multiple intelligences)
There are at least eight different kinds of intelligence (...and rising): logical, linguistic, spatial, musical, kinesthetic, naturalist, intrapersonal and interpersonal

Gardner, Howard (1993). Frames of mind: The theory of multiple intelligences. New York: Basic Books

Peter Salovey and John D. Mayer – **Emotional Intelligence**

The ability to recognize, monitor one's own and others' emotions, to discriminate among them and to use this information to guide one's thinking and actions.

Salovey, P. & Mayer, J.D. (1990) "Emotional intelligence" *Imagination, Cognition, and Personality*, 9, 185-211



Definition and classification of emotions

Definition of Emotions

Emotions = every temporary state of mind, feeling or emotional state evoked by experiencing different sensations.

Nakamura, A.: Kanjo hyogen jiten (Dictionary of Emotive Expressions) (in Japanese), Tokyodo Publishing, Tokyo (1993)

Emotive utterances = every utterance in which the speaker in question is emotionally involved, and in which this involvement is linguistically expressed by means of intonation or by the use of performative expressions.

Beijer, F.: The syntax and pragmatics of exclamations and other expressive/emotional utterances. Working Papers in Linguistics 2, The Department of English in Lund. (2002)

Nakamura's classification of emotions (after a thorough study in the Japanese):

10 types:

- | | |
|--|---|
| 1. 喜 <i>ki</i> / <i>yorokobi</i> [<u>joy, delight</u>] | 6. 好 <i>kou</i> / <i>suki</i> [<u>liking, fondness</u>] |
| 2. 怒 <i>do</i> / <i>ikari</i> [<u>anger</u>] | 7. 厭 <i>en</i> / <i>iya</i> [<u>dislike, detestation</u>] |
| 3. 哀 <i>ai</i> / <i>aware</i> [<u>sorrow, sadness</u>] | 8. 昂 <i>kou</i> / <i>takaburi</i> [<u>excitement</u>] |
| 4. 怖 <i>fu</i> / <i>kowagari</i> [<u>fear</u>] | 9. 安 <i>an</i> / <i>yasuragi</i> [<u>relief</u>] |
| 5. 恥 <i>chi</i> / <i>haji</i> [<u>shame, shyness, bashfulness</u>] | 10. 驚 <i>kyou</i> / <i>odoroki</i> [<u>surprise, amazement</u>] |



Our approach

Recognition of emotions:

- Voice
- Facial expressions
- Gestures
- **Language**

	textual data	voice and visual data
gathering	easy	laborious
available data	many corpora plus Web	only prepared for the particular research
processing	fast	slow and heavy
semantics	OK.!	NO!



In language there are:

1. Emotive expressions. Parts of speech, that in emotive sentences describe emotional states.

A. Nakamura, *Kanjō hyōgen jiten* (Dictionary of Emotive Expressions), Tokyodo Publishing, Tokyo (1993)

Examples: nouns: *aijou* (love); verbs: *kanashimu* (feel sad); adjectives: *ureshii* (happy)

2. Emotive elements. Indicating that emotions have been conveyed, but not detailing what specific emotions there are. The same emotive element can express different emotions depending on context.

M. Ptaszyński, *Moeru gengo - Intānetto kei-jiban no ue no nihongo kaiwa ni okeru kanjōhyōgen no kōzō to kigōrontekikinō no bunseki* – "2channeru, denshikeijiban o rei toshite" – (Boisterous language. Analysis of structures and semiotic functions of emotive expressions in conversation on Japanese Internet bulletin board forum - 2channel -), UAM, Poznań (2006)

Examples:

interjections/exclamations: *sugee* (great!); mimetics: *wakuwaku* (heart pounding);

vulgar language: *-yagaru* ("fu**ing do sth" – a vulgarisation of a verb)

Emotive Elements / Expressions Analysis System (ML-Ask)

emotive
expr.
DB

emotive
elements
DB

nouns

愛情 *ajjou* (love)
安心 *anshin* (relief)
恐怖 *kyofu* (fear)

verbs

喜ぶ *yorokobu* (be glad)
悲しむ *kanashimu* (feel sad)
むかつく *mukatsuku* (get angry)

phrases / idioms

虫酸が走る *mushizu ga hashiru* (give one the creeps)
心が解ける *kokoro ga tokeru* (one's heart is melting in relief)
歓天喜地 *kantenkichi* (delight larger than Haven and Earth)

adjectives

嬉しい *ureshii* (happy)
悔しい *kuyashii* (mortifying)
怖い *kowai* (scary)

exclamatives

すげえ *sugee* (great!)
うおお *wooo* (whoa!)

mimetics (*gitaigo*)

ワクワク *wakuwaku* (heart pounding)
ドキドキ *dokidoki* (go pit-a-pat)

vulgarities

-やがる *-yagaru* (fu**ing do sth)
くそ *kuso* (shit)
馬鹿 *baka* (stupid)

hypocoristics

-ちゃん *-chan* (name suffix)

textual representations of voice modulation and body language (emoticons)

“!” , “??” , “...” , (T_T) , (-_-;) , _|_|_O



An example of analysis

この本さー、すげー やばかった よ。まじ**怖**すぎ。
Kono hon saa, sugee yabakatta yo. Maji kowa sugi.
That book, ya know, it was a killer. It was just too scary.

emotive elements:

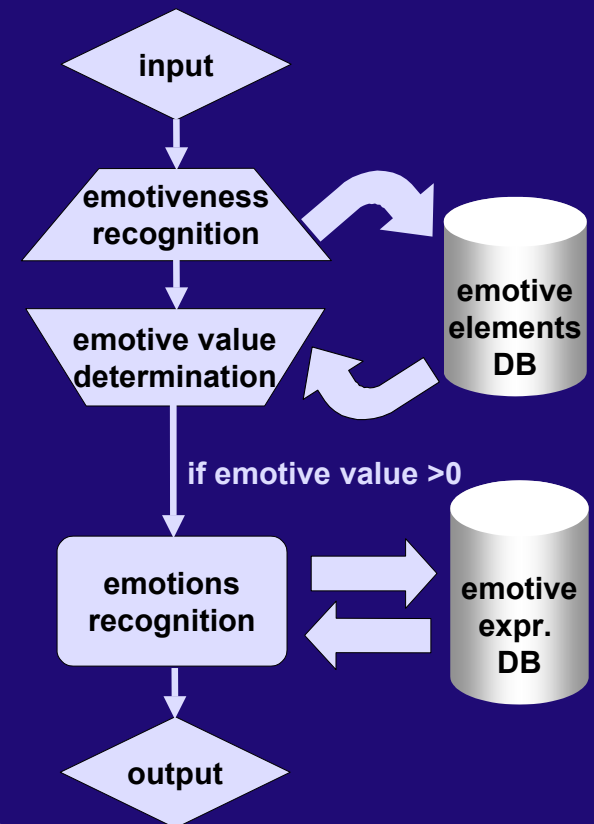
さー, すげー, やばい, -よ, まじ

emotive value = 5

emotive expressions:

怖い

System Flowchart



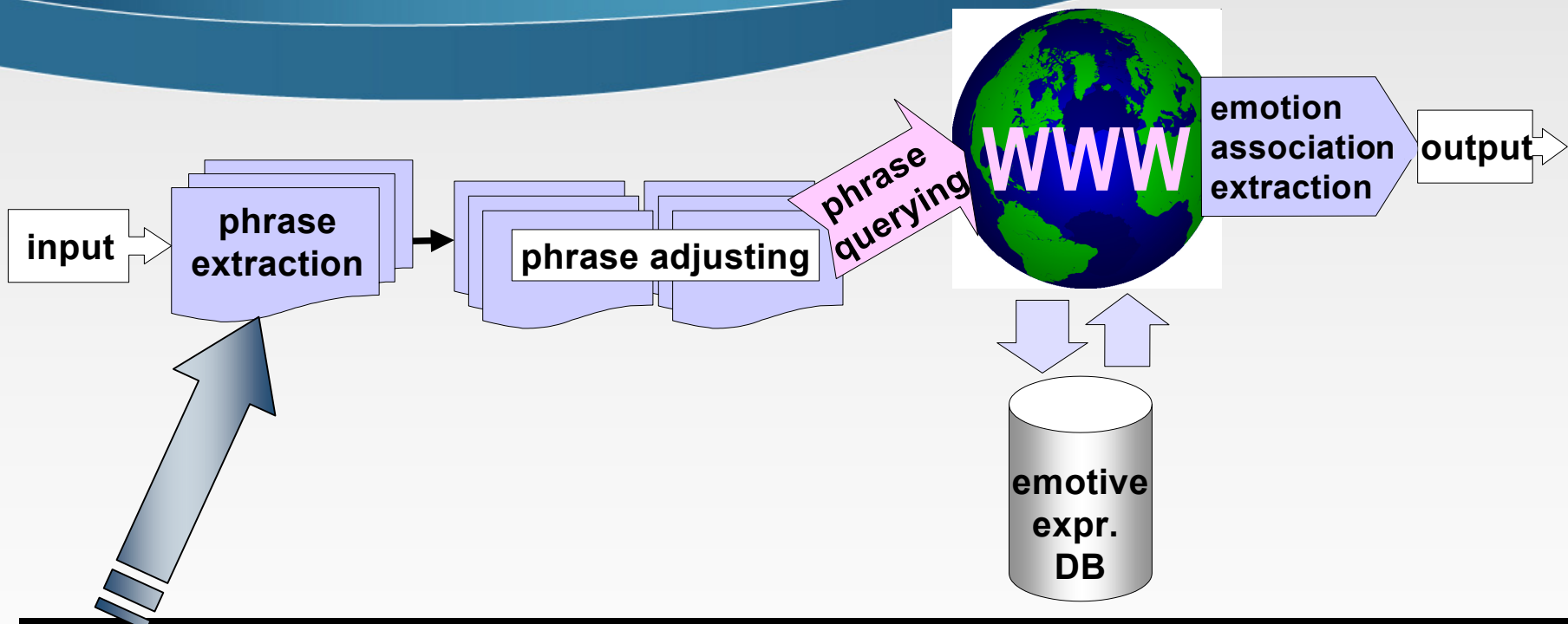
Applying Shi's Web mining technique for extracting emotive associations from the Web* to improve the extraction of the emotion types in ML-Ask.

Shi, Wenhan. "Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet." *FIT2008*, pp. 267-268, 2008.

- * A) ML-Ask with Shi's technique activated only for utterances where baseline system didn't manage;
- B) ML-Ask with Shi's technique instead of the usual emotion extraction method;

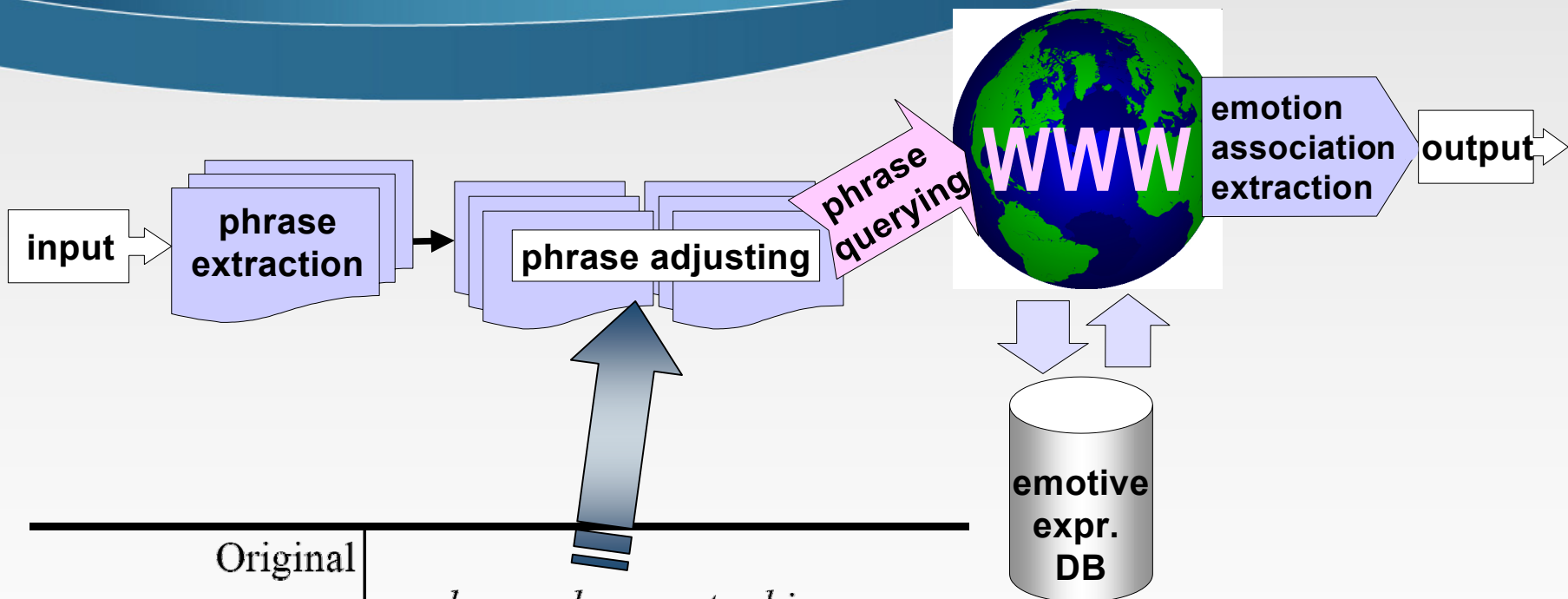


Web Mining Technique



Original utterance	<i>Aa, pasokon ga kowarete shimatta...</i> (Darn, the PC has broken...)	
longest n-gram (here: hexagram)	<i>Aa pasokon ga kowareru te shimau</i>	
	[interjection] [noun] [particle] [verb] [verb connector] [perfect form]	
pentagram	<i>pasokon ga koware te shimau</i>	
tetragram	<i>Aa, pasokon ga kowareru</i>	
trigrams	<i>pasokon ga kowareru</i>	<i>koware te shimau</i>

Web Mining Technique

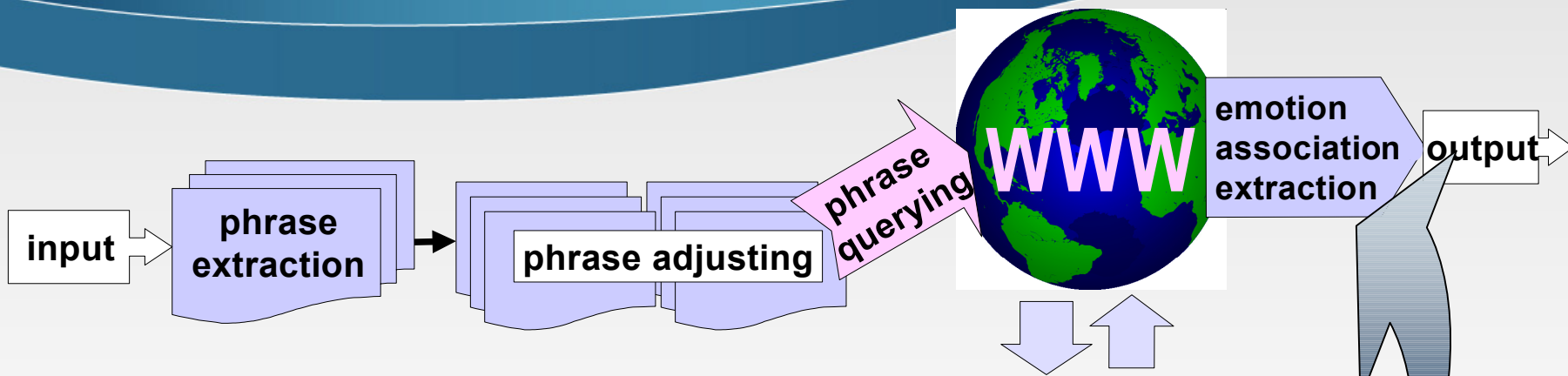


	Original n-gram	
	<i>pasokon ga koware te shimau</i>	
n-gram phrase adjusting (morpheme modification)	/ -te /	<i>pasokon ga koware te shima-tte</i>
	/ -to /	<i>pasokon ga koware te shimau to</i>
	/ -node /	<i>pasokon ga koware te shimau node</i>
	/ -kara /	<i>pasokon ga koware te shimau kara</i>

morphemes of causality

-te, -to, -node, -kara, -ba, -nowa, -noga, -kotowa, -kotoga

Web Mining Technique



Sentence: *Aa, pasokon ga kowarete shimatta...*(Darn, the PC has broken...)

Extracted emotion type	Type extracted / all extracted types	Ratio
[fear]	28/133	0.21052631578947
[sorrow, sadness]	26/133	0.19548872180451
[dislike, detestation]	16/133	0.12030075187969
[liking, fondness]	14/133	0.105263157894737
[relief]	12/133	0.090225563909774
[excitement]	11/133	0.082706766917293
[joy, delight]	10/133	0.075187969924812
[surprise, amazement]	9/133	0.067669172932330
[anger]	5/133	0.037593984962406
[shame, shyness, bashfulness]	2/133	0.015037593984962

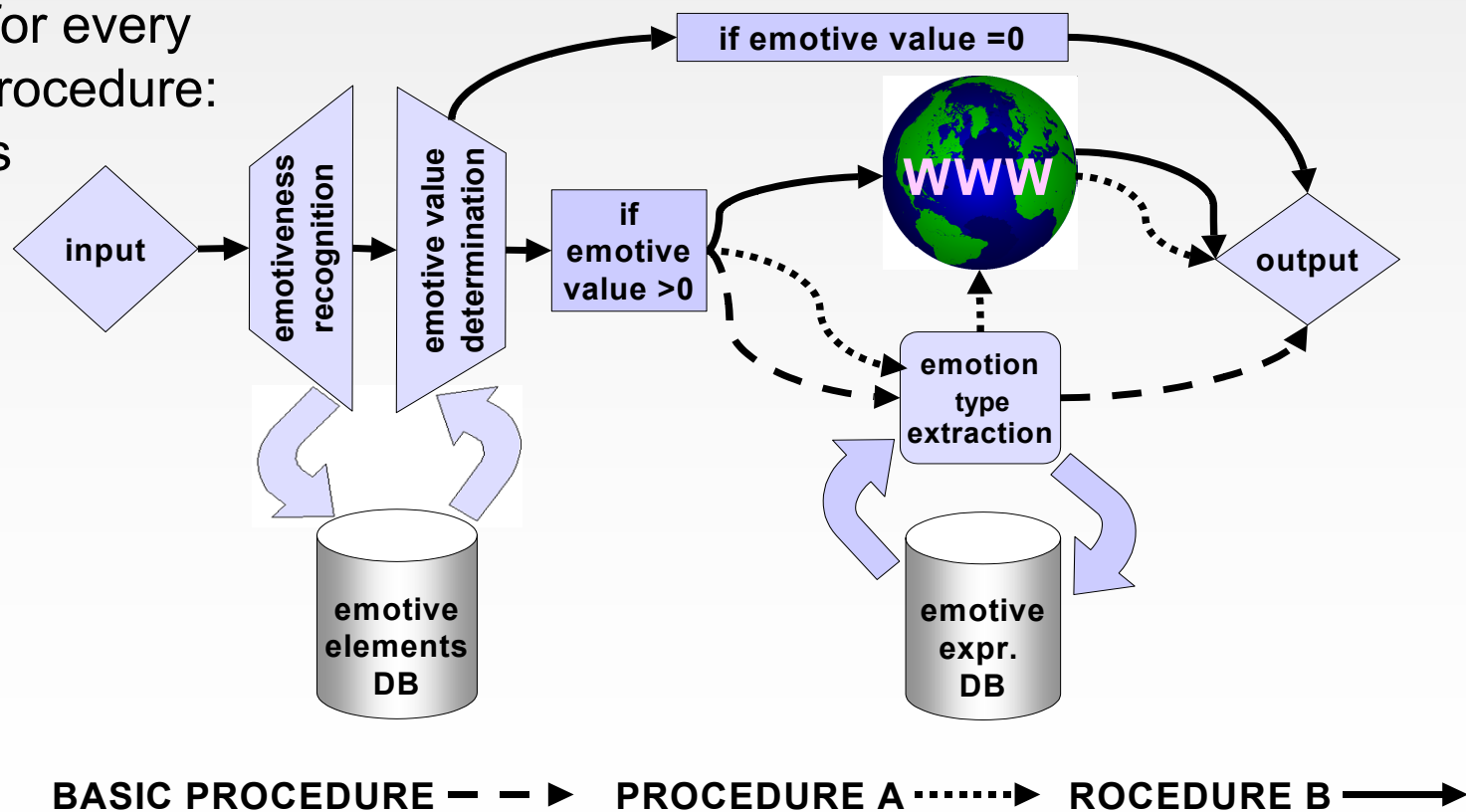
Variant 2
3 highest results

Variant 1
all emotions



Web Mining Technique

- Two kinds of procedure (A – support; B – instead)
- Two variants for every Web mining procedure:
 - all emotions
 - 3-highest





I think
therefore I
am.

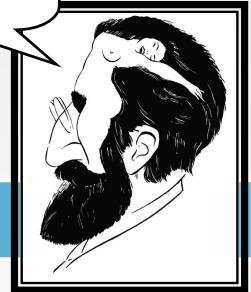
Two perspectives towards emotions:

“Cartesian“

- Descartes about emotions:

Only the one who expresses emotions knows exactly what they are.

You wanna
talk about
it?



“Freudian“

- Freud about emotions:

The one who expresses emotions know nothing about them – better ask the third person.

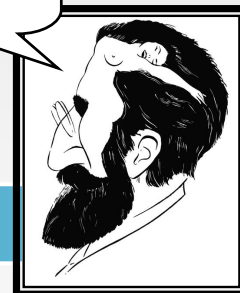


I think
therefore I
am.

Two perspectives towards emotions:

“Cartesian“

“Froedlian“



You wanna
talk about
it?

- Descartes about emotions:

Only the one who expresses emotions knows about them. Emotions are.

**Separately they were
both wrong...**

- Freud about emotions:

Emotions are expressed by the person. Talking about them – better than not talking about them.

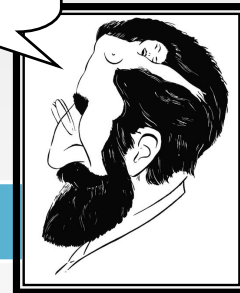


I think
therefore I
am.

Two perspectives towards emotions:

“Cartesian“

“Freudian“



You wanna
talk about
it?

- Descartes about emotions:

Only the one who expresses emotions knows exactly what they are.

Truth:

Neither the speaker (first person) nor the listener (third person) can always know the emotional states (of the speaker) for sure.

- Freud about emotions:

People who express emotions about them – better person.

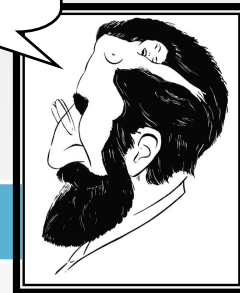


I think
therefore I
am.

Two perspectives towards emotions:

“Cartesian“

“Freudian“



You wanna
talk about
it?

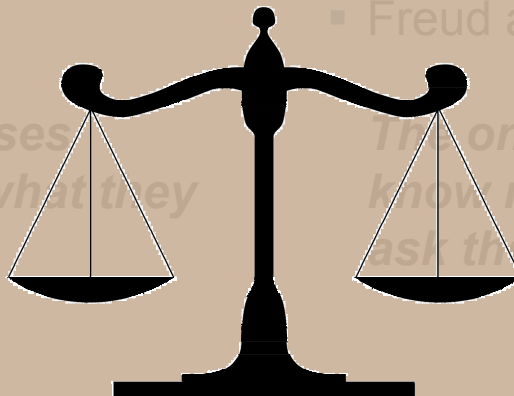
- Descartes about emotions:

Only the one who expresses emotions knows exactly what they are.

- Freud about emotions:

The one who expresses emotions knows nothing about them – better ask the third person.

What we need is:



Balance between them

“Cartesian“

“Freudian“

Collection of utterances (90) is tagged emotively by:

- authors of the utterances.
- a third party evaluators (8-12 people).

We compare system’s output to it to calculate:

- the level of recognition of emotions according to the authors of utterances.
- the level of agreement with general human commonsense about what emotions were conveyed in utterances.

Evaluation

Recognition ~~“Cartesian”~~

~~“Frenchan”~~ Commonsense

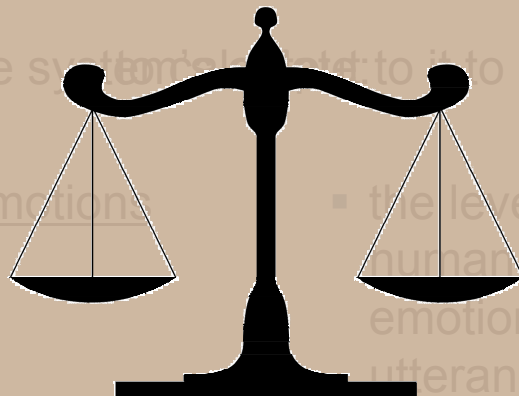
Collection of utterances (90) is tagged emotively by:

- authors of the utterances.

We

- the level of recognition of emotions according to the authors of utterances.

Ideal result:



Both high and balanced

by third party evaluators.

- the level of agreement with general human commonsense about what emotions were conveyed in utterances.

Recognition

Commonsense

Emotiveness (emotive/non-emotive)

- **.83** (balanced mean separately for emotive and non-emotive)
- ≈
- **86%** (approx. agreement with 3rd party evaluators)

Emotive value (0-5)*

- **.53**
- ≠
- **82%**

* with applied condition of almost-perfect match (emotive value differs between two compared subjects, e.g. author of an utterance vs the system, two evaluators, or evaluator vs system, by only ± 1 emotive point.)

Recognition

Commonsense

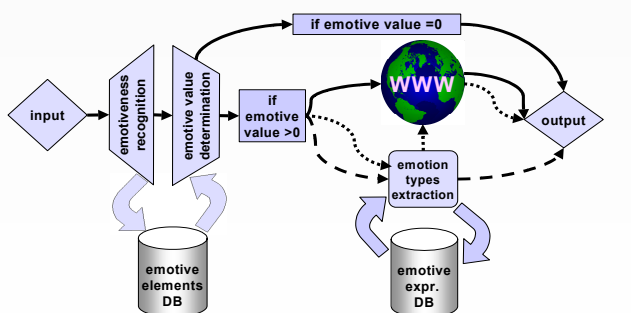
Emotion type extraction (base line)

- **.45** (balanced mean for all emotions including non-emotive)

≈

- **45%** (approx. agreement with 3rd party evaluators; two conditions:

1. at least one emotion type from the emotion type list by humans
2. system's result = majority on the emotion type list by humans)



BASIC PROCEDURE - - - - -> PROCEDURE A - - - - -> PROCEDURE B - - - - ->

Evaluation

Recognition

Commonsense

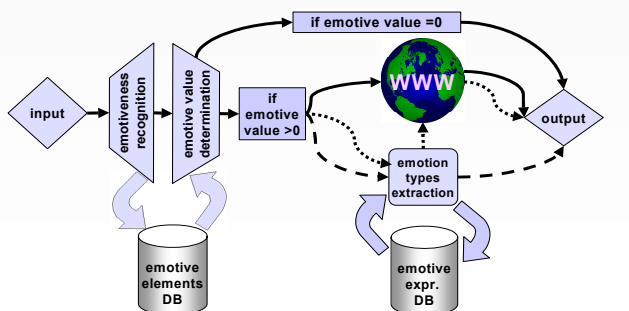
Emotion type extraction

Procedure A

Variant 1

■ .53

■ 60%



Variant 2

■ .54

■ 67%

BASIC PROCEDURE - - - - -> PROCEDURE A - - - - -> PROCEDURE B - - - - ->



Evaluation

Recognition

Commonsense

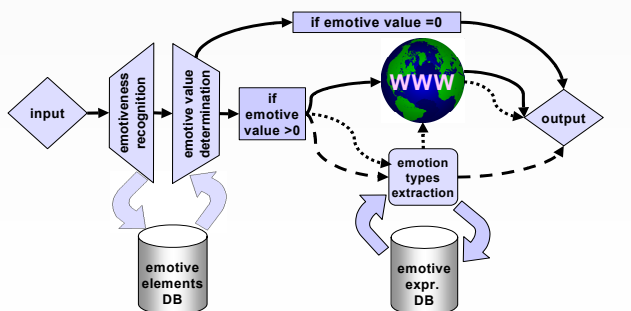
Emotion type extraction

Procedure B

Variant 1

■ **.53**

■ **50%**



Variant 2

■ **.61**

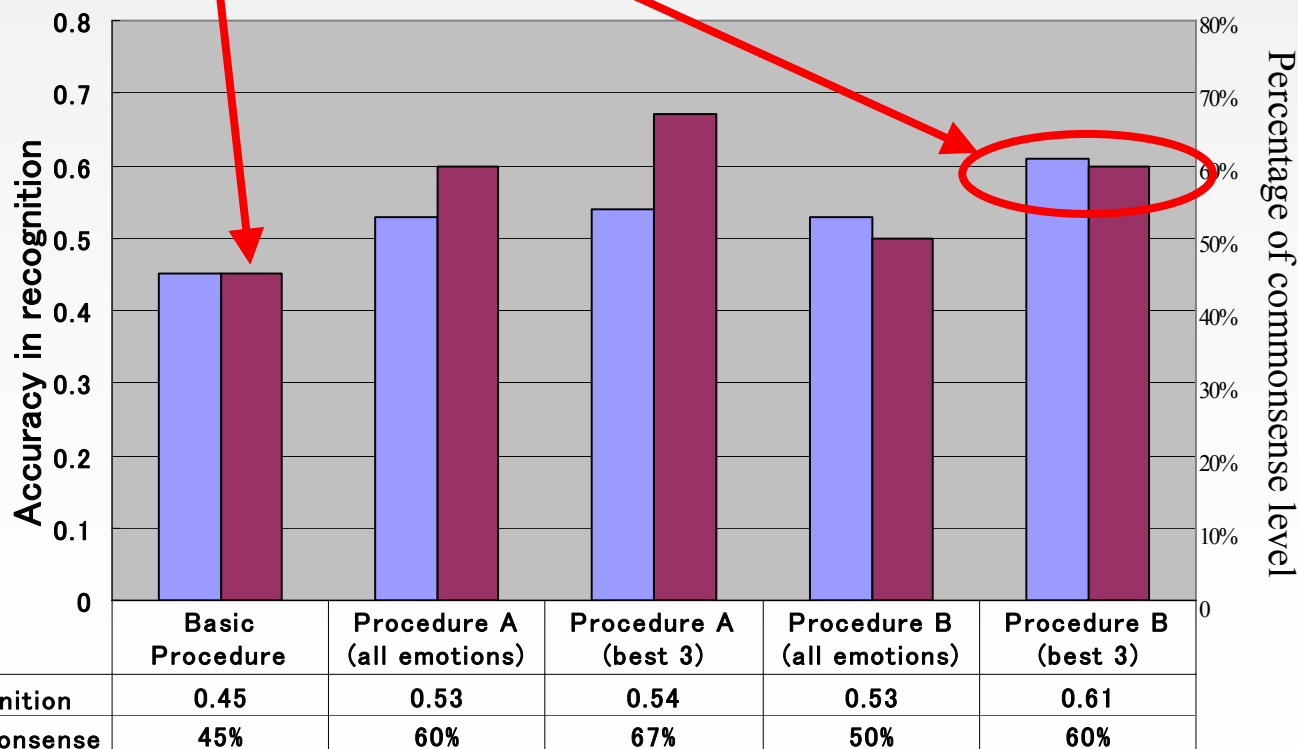
■ **60%**

BASIC PROCEDURE - - -> PROCEDURE A - - - - -> PROCEDURE B - - - - ->



Results

- ML-Ask's performance was enhanced by Web mining.
- Balanced procedure was specified



Conclusions

- 👍 In every case the Web mining improved emotion types determination.
- 👍 It was more effective to keep only 3 highest results of the emotion types extracted from the Web.
- 👍 The balanced method was determined.
- 👍 The system could be useful in such empirical methods as:
 - Extraction of Information (about emotions) from text
 - Tagging (the whole corpora with emotive tags)
 - Word Sense Disambiguation (in emotive/non-emotive speech)
 - Agent Modeling (determining a choice of action patterns according to user's emotions)
 - Education (Teaching foreigners natural [Japanese] language)
 - Cultural studies (Creating multi-language global version of the system could answer the question - How people in different cultures express emotions through language)



➤ **Still needed:**

- More accurate phrase extraction from an utterance
- More people for evaluation
- Larger corpus for evaluation (whole conversations)

➤ **Russell's two-dimensional model of emotions is likely to help in determination of emotion types (ex. What emotions go with “!” ?).**

➤ **Applying a Web page indexing (like HyperEstraiier, etc.) should speed up the Web mining process (now a few minutes to even a dozen or so) and make the system applicable to real time affect analysis.**

➤ **Next application → emotionally conscious conversational agent (choosing the type of action depending on user's emotion – at present – experiments in progress)**

➤ **Final application → multipurpose tool for computing emotive function of language**

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



Thank you for your kind attention!



北海道大学
HOKKAIDO UNIVERSITY



荒木研究室
ARAKI LABORATORY