

Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki

Ideas for Using Large-Scale Corpora to Improve Verification of Emotion Appropriateness in Japanese

ABSTRACT

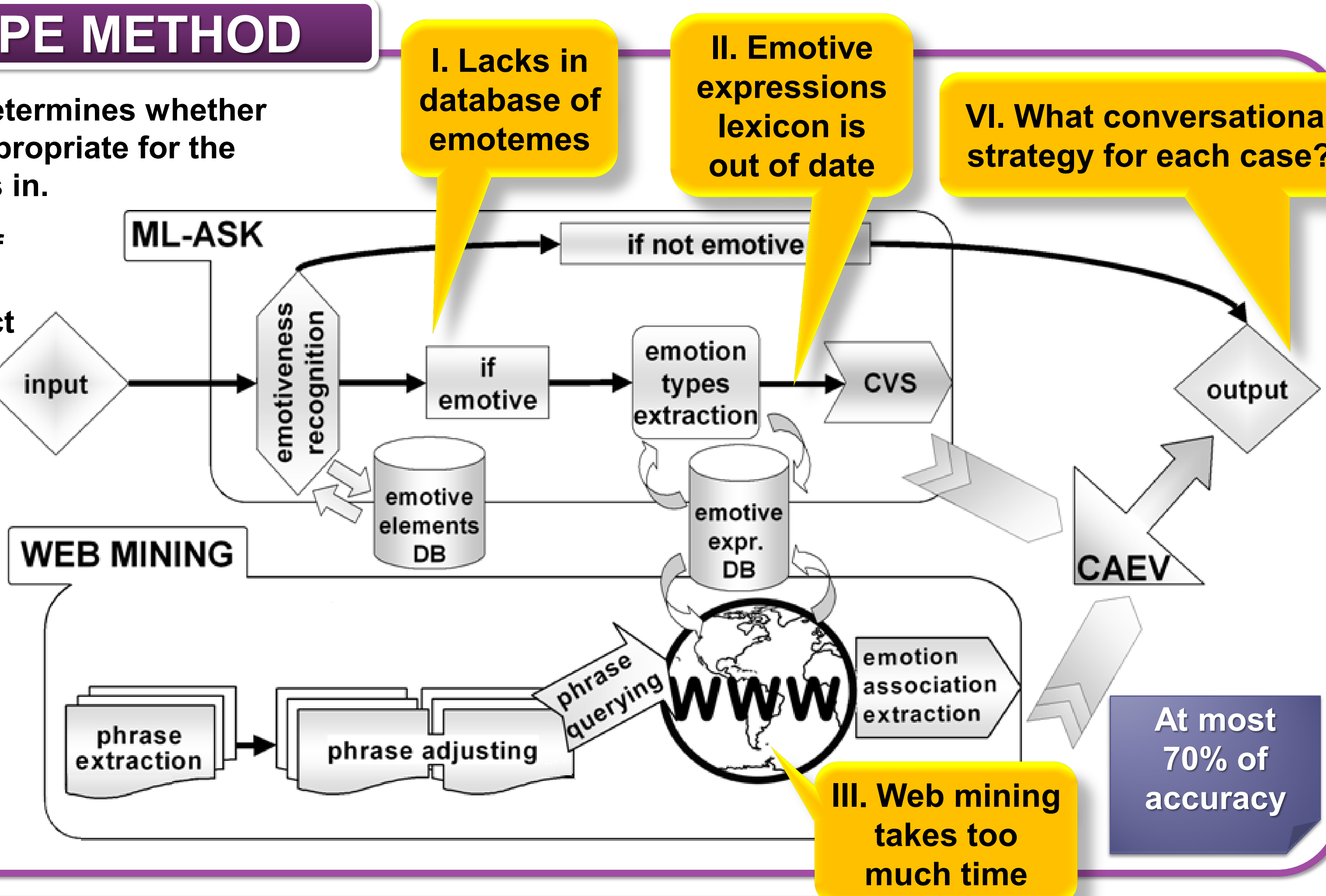
We present a set of ideas for improving the verification of emotion appropriateness in Japanese language. Emotion appropriateness verification is a new method for discovering not only what are the emotions conveyed by a user in an utterance also whether they are appropriate for the context they are used in. We present ideas to improve this method with the use of several corpora. The corpora we plan to use are Amazon reviews, Web as corpus and two corpora of natural conversations to improve the method and provide conversational strategies for implementation of the method into a conversational agent.

PROTOTYPE METHOD

The method [1] determines whether the emotion is appropriate for the context it appears in.

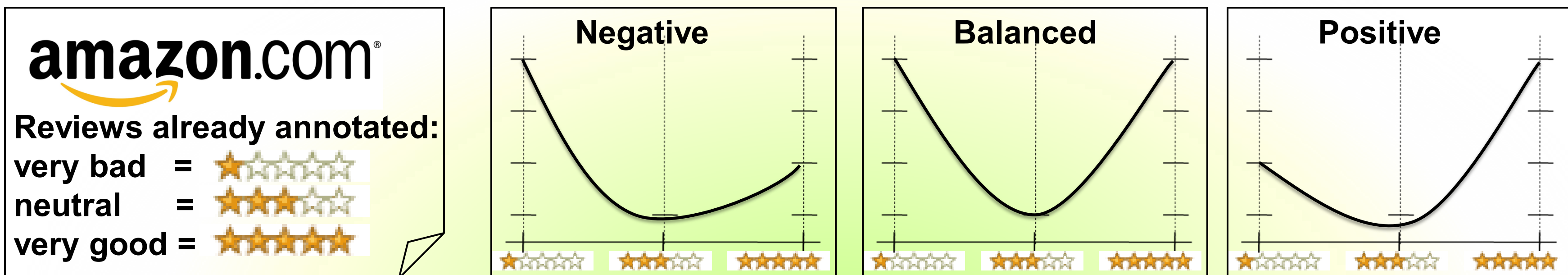
For recognition of emotions it uses Ptaszynski's affect analysis and annotation system, ML-Ask [2],

To verify the contextual appropriateness of those states it uses Shi's Web mining technique [3] for gathering emotive common sense from the Internet.



I. OPTIMIZING DATABASE OF EMOTEMES

1. REFINING EMOTEMES: Extracting expressive words appearing most often in 1* and 5* reviews [4].



CROSS-REFERENCE:

Reconfirming the rest of candidates:

1) Extracting example-sentences (30) for each unconfirmed emoteme form 3 different corpora

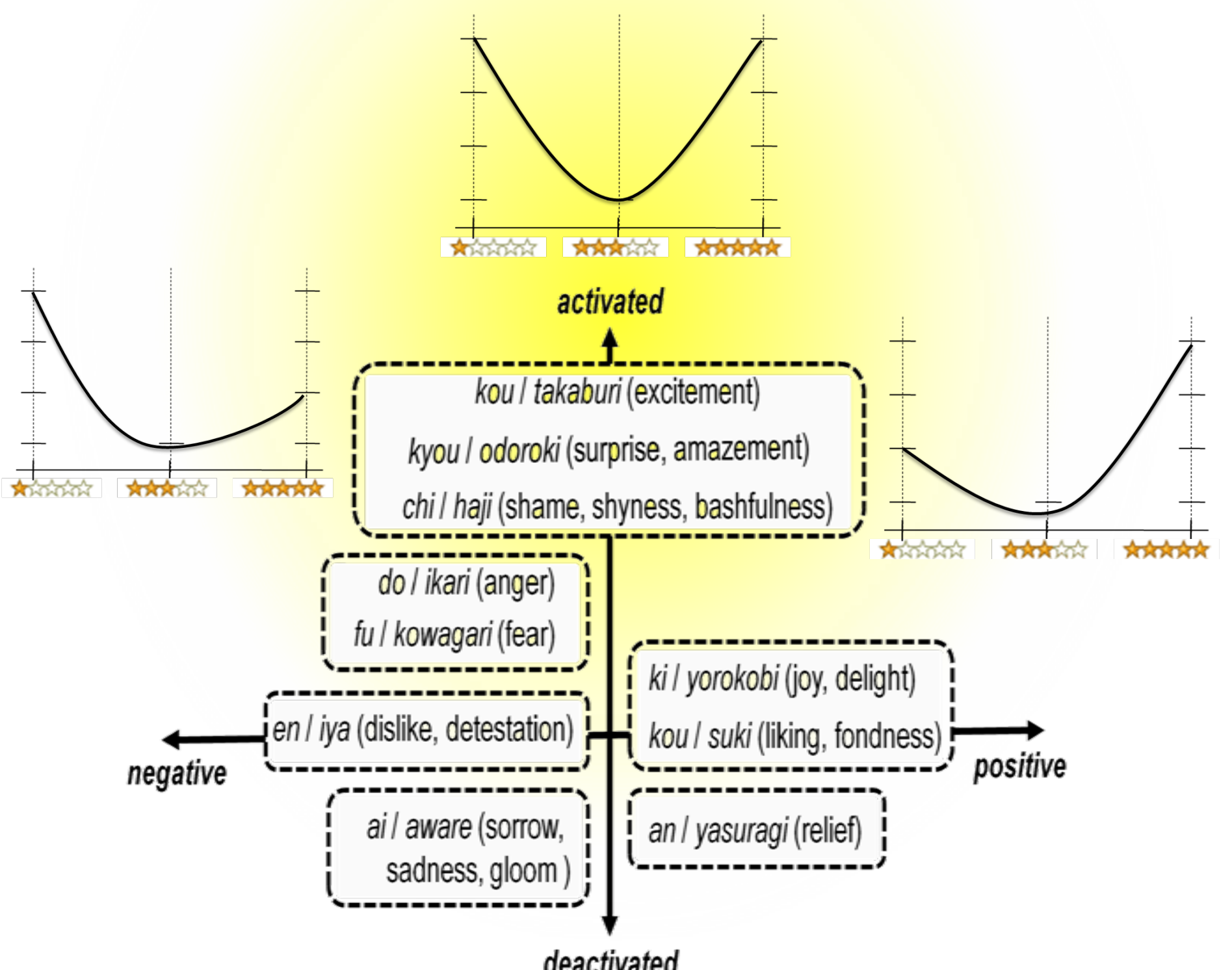
Sentence 01 ... Sentence 10
Corpus A [5] Corpus B [6] Corpus C [7]

2) Questionnaire for each sentence:

"Do you think author of this utterance was conveying any emotions, or was he/she neutral?"

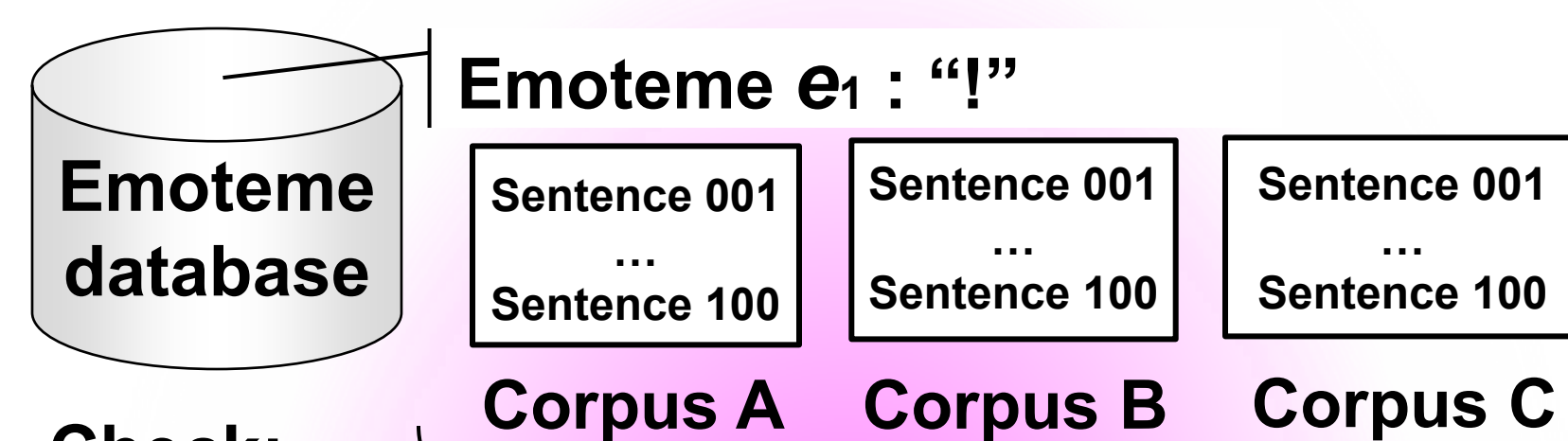
2. DISAMBIGUATING EMOTEMES:

1) Limiting the list of possible emotion types/word:
-Use Amazon annotations to verify polarity tendency of a word ("J-shape" or "reversed J-shape").
-Use mapping of Nakamura's [8] emotion types on Russell's [9] 2-D model of affect, like in [1].



2) Statistical disambiguation:

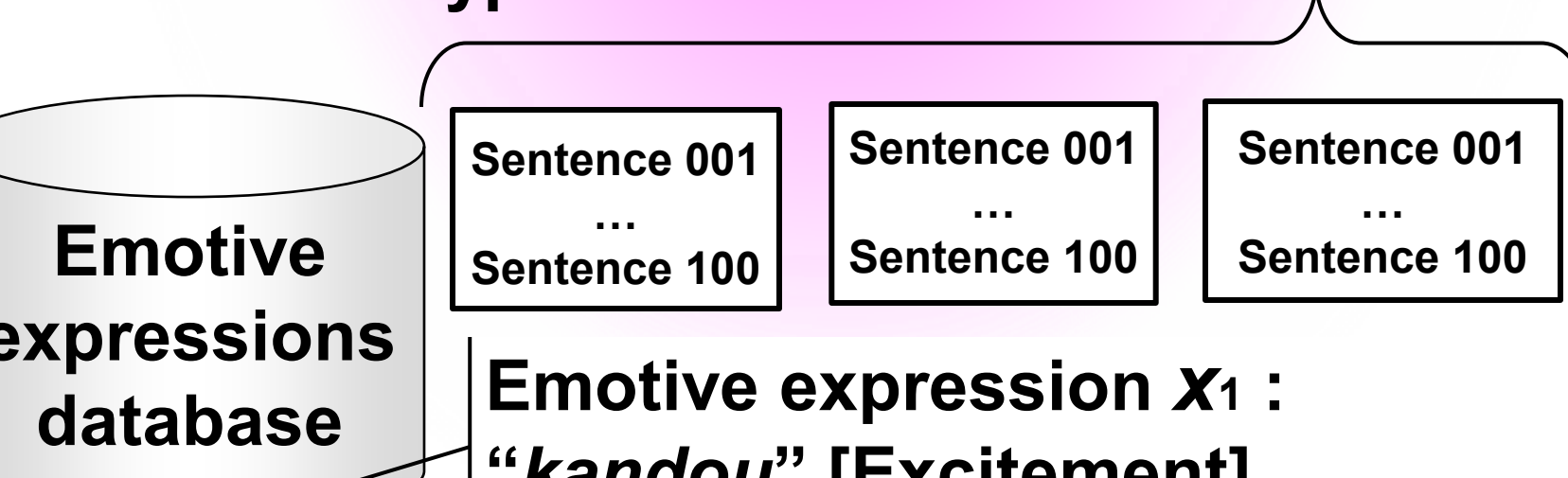
-For every emoteme and emotive expression take 100 sentences (or all applicable) from every corpus.



-Check:

1. Most frequent emotion types for an emoteme and:

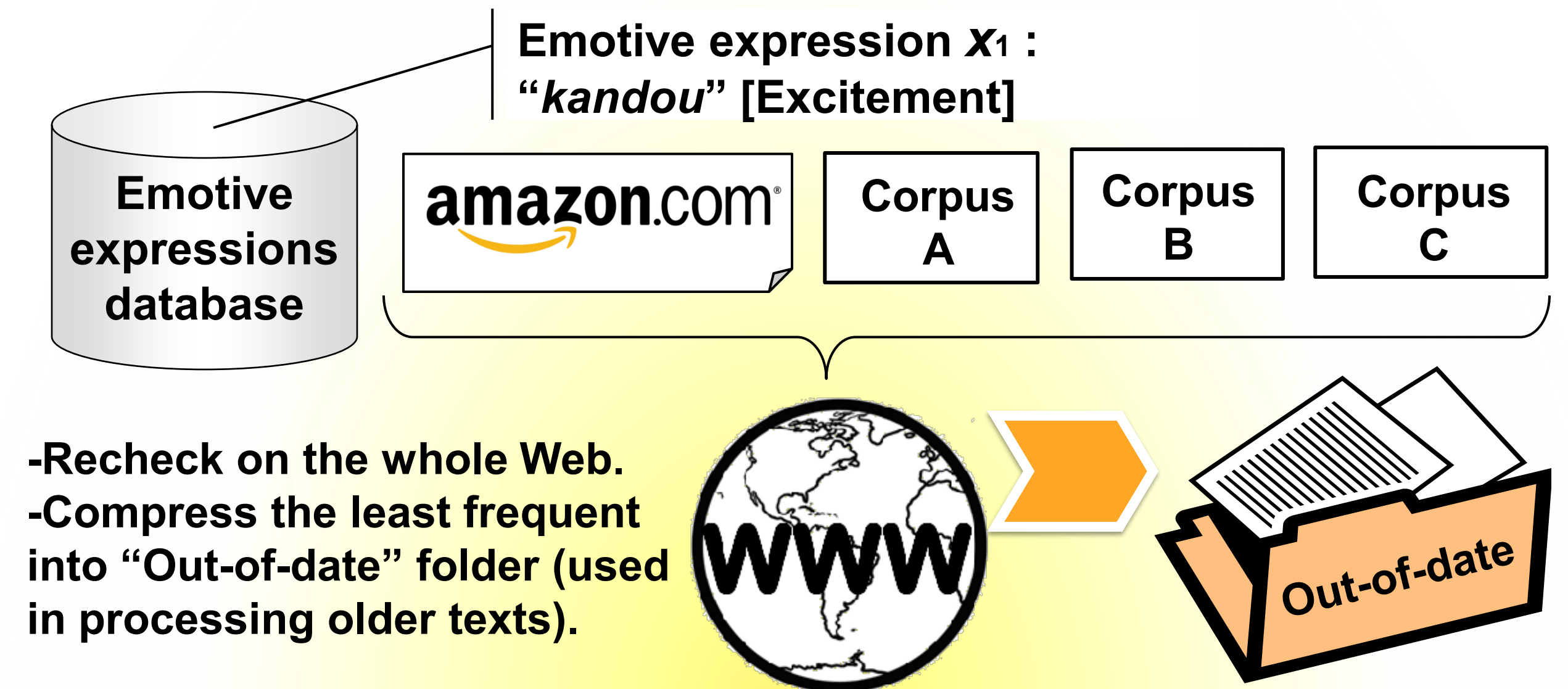
2. Most frequent emotemes for an emotion type



II. UPDATING DATABASE OF EMOTIVE EXPRESSIONS

1. REFINING EMOTIVE EXPRESSIONS:

-Check the concordance (hit rate) of every emotive expression from the data-base in Amazon and all three corpora.



2. EXPANDING EMOTIVE EXPRESSION DATABASE:

-Extract syntactical patterns of emotive expressions from emotive sentences:

Kyou wa nante kimochi ii hi da na !
today:THEM [e1] [x1:joy] day:OBJ [e2] [e3]
Today is such a nice day!
Pattern: THEM [e1] @ OBJ [e2] [e3] (x2 = @)
Sore wa nante osoroshii koto dana !
x2 = osoroshii [fear]

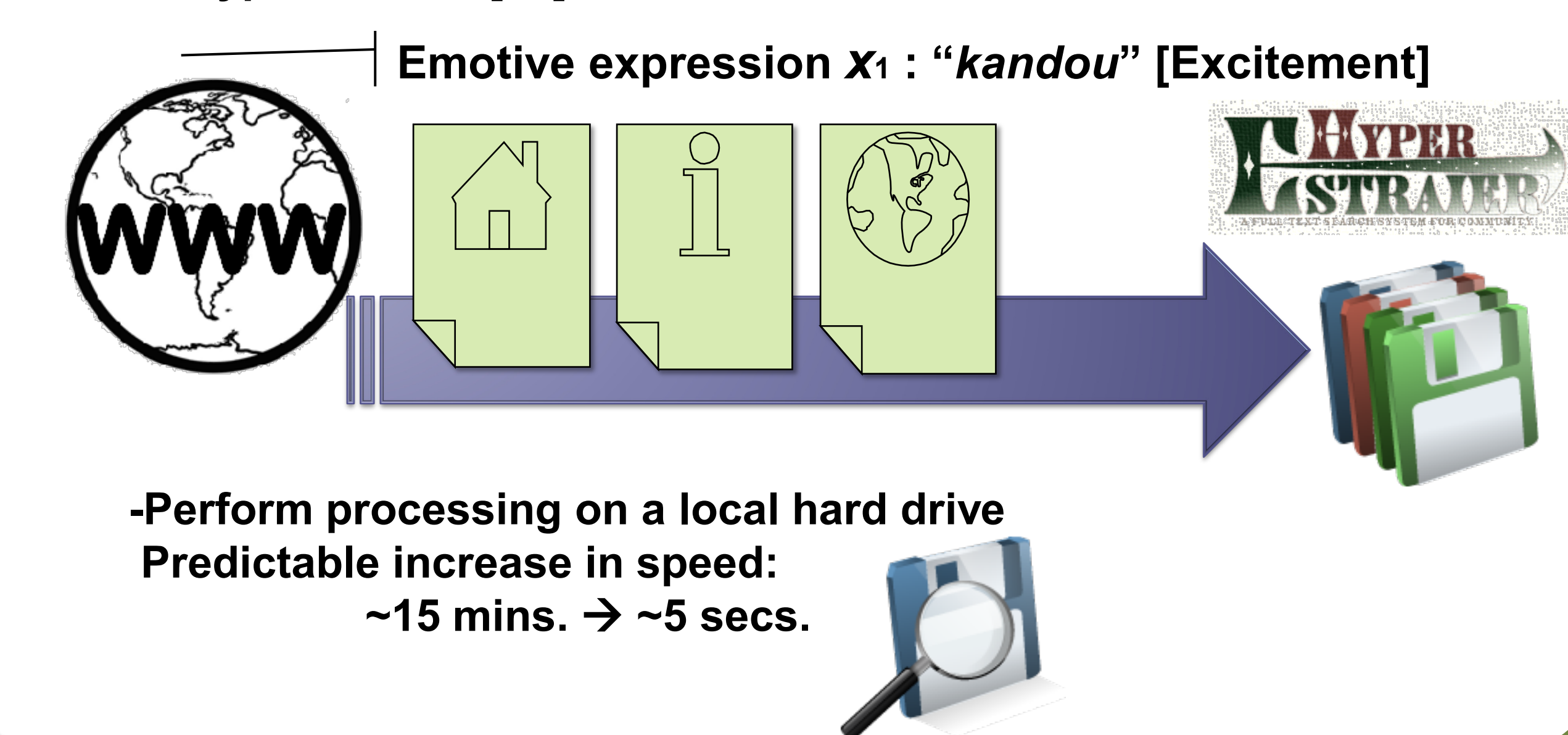
-Questionnaire for sentences with each new emotive expression candidate:

"Do you agree that this sentence expresses [emotion type]?"

III. IMPROVING WEB MINING

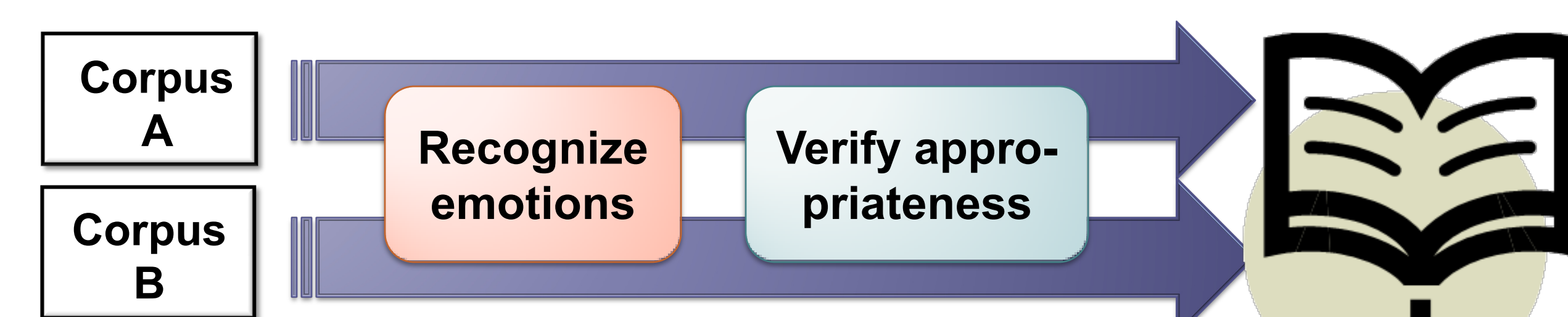
1. FROM BROWSER WEB MINING TO A STAND ALONE SYSTEM:

-Gather large number of pages from the Web to a hard drive and index with HyperEstrailer [10].



IV. EXTRACTING CONVERSATIONAL STRATEGIES

1. ANALYSE CONVERSATION CORPORA WITH THE METHOD



2. FIND CONVERSATIONAL STRATEGIES

-Use the set of conversational strategies for the Japanese language [11].
-Extract sentence patterns with valuable conversational strategies

3. IMPLEMENT IN CONVERSATIONAL AGENT

CONCLUSIONS

- We presented a set of ideas for improving a new method for:
1) discovering what are the emotions conveyed by a user in an utterance
2) whether they are appropriate for the context they are used in.
-The planned improvements include using of different, large-scale corpora, like Amazon reviews, Web as Corpus, or two conversation corpora.
-We also plan to extract conversational strategies for further implementation of the system into a conversational agent.

REFERENCES

1. M. Ptaszynski, P. Dybala, W. Shi, R. Rzepka and K. Araki, "Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States." In Proceedings of IJCAI-09, Pasadena, CA, USA, pp. 1469-1474, 2009.
2. M. Ptaszynski, P. Dybala, R. Rzepka and K. Araki, "Affecting Corpora: Experiments with Automatic Affect Annotation System -A Case Study of the 2channel Forum-," In Proc. of PACLING-09, Japan, 2009.
3. W. Shi, R. Rzepka and K. Araki, "Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet," FIT-08, pp. 267-268, 2008.
4. C. Potts and F. Schwarz, "Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora." Ms., UMass Amherst, 2008.
5. M. Usami (Ed.), "BTS ni yoru nihongo hanshikotoba kopasu1 (hatsu tamen, yuujiin; zatsudan, touron, sasoi) [Conversation corpus of spoken Japanese using the Basic Transcription System (first meeting, friend's conversation, small talk, discussion, invitation)]," Tokyo University of Foreign Studies, Tokyo, Japan, 2007.
6. http://tell.fll.purdue.edu/chakoshi/public.html
7. K. Yoshihira, T. Takeda, and S. Sekine, "KWIC system for Web Documents" (in Japanese), In Proceedings of NLP-2004, pp. 137-139, 2004.
8. A. Nakamura, Kanjo hyogen jiten, Tokyodo, 1993.
9. J. A. Russell, A circumplex model of affect, Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980.
10. http://hyperestraier.sourceforge.net/index.html
11. Y. Nakai, M. Ooba, M. Doi, "Danwa reberu de no kaiwa kyoiku ni okeru shido komoku no teian - danwa / kaiwa bunseki teki apurochi no kanten kara mita danwa gino no komoku," Sekai no nihongo kyoiku, 14, 2004.