Bridging quantitative and qualitative methods for social sciences
using text mining techniques

TAKMI (Text Analysis and Knowledge Mining) **and Sentiment Analysis**

April 28, 2006

Tetsuya Nasukawa
IBM Research, Tokyo Research Laboratory

Outline

- TAKMI
  - ✓ Overview
  - ✓ Application Examples
    - Customer Contact Records
    - Technical Documents
      - Patent
    - Medical Documents
      - Medline
  - ✓ Demo
- Sentiment Analysis
  - ✓ Application Overview
    - Reputation Mining
  - ✓ Technical Overview
  - ✓ Progress
- Conversation Mining
  - ✓ Overview
  - ✓ Demo

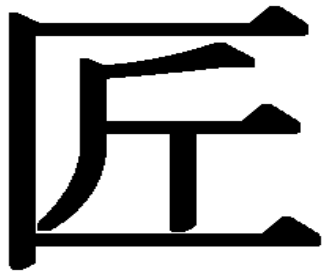
TAKMI

TAKMI
Overview

IBM TAKMI

TAKMI stands for **T**ext **A**nalysis and **K**nowledge **Mi**ning.
It also has Japanese meanings.

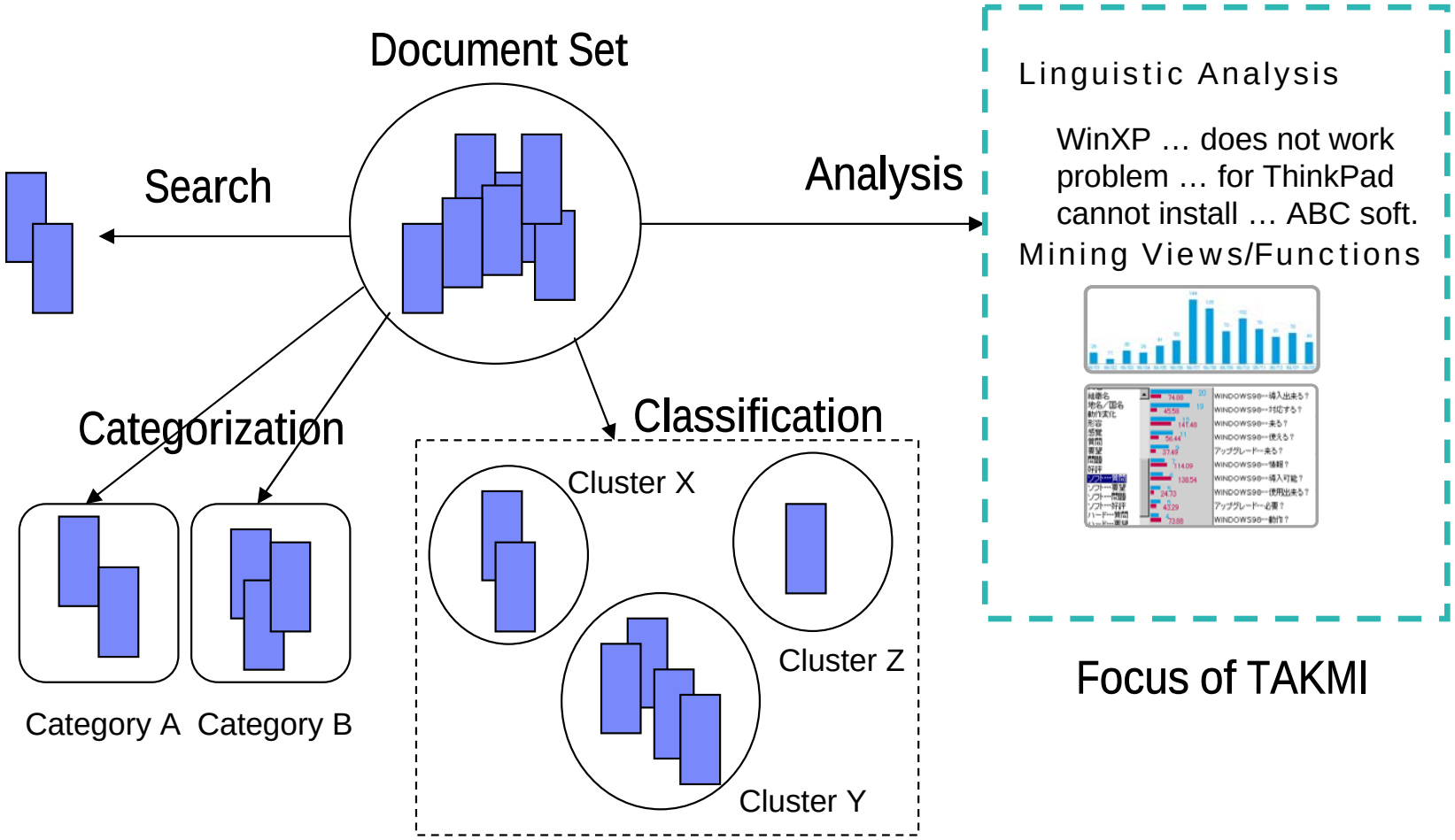
= skillful, clever



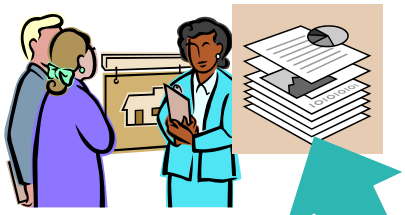
= craftsman, artisan

What is TAKMI?

TAKMI focuses on analyzing what a large set of documents indicates as a whole.



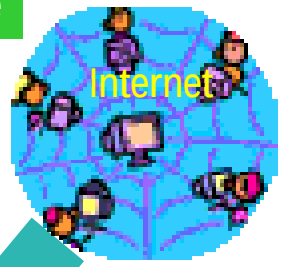
Using the Customers Voices' with TAKMI



Marketing and Customer Intelligence

Do not have enough information to select customers to contact for a campaign

- Extracting potential customers from contact logs



Want to know users' reactions

- Extracting reputation from users about products/services

Contact Center

Sales Productivity Improvement

Want to enhance operators' skill

- Use of good contact patterns of good operators in education course

Contact Center Business Process Optimization

Want to reduce the number of calls

- FAQ creation assistance

Cannot read too many call logs

- Creation of monthly reports
- Early detection of problems

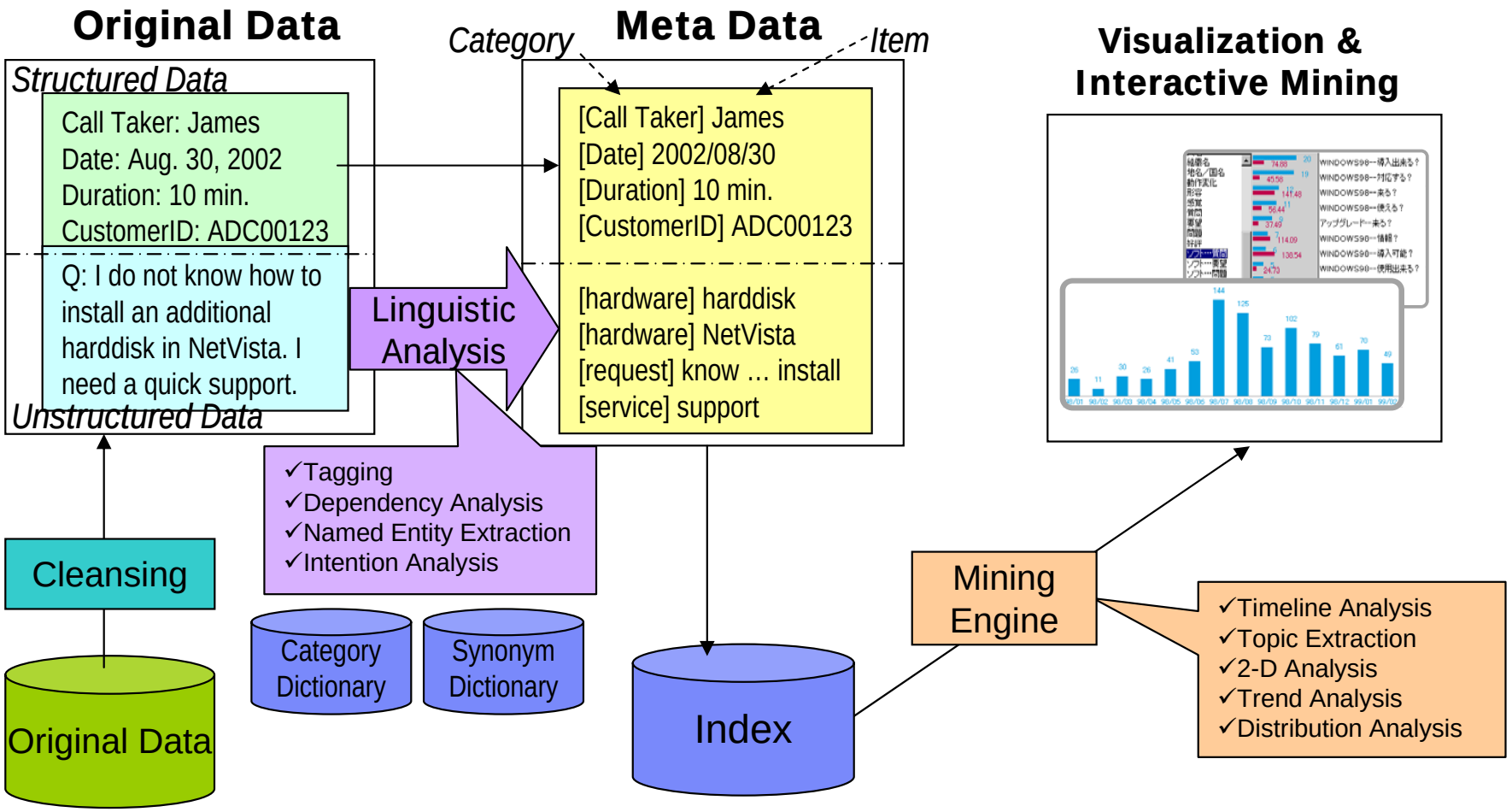
Call Log DB
(Inbound/Outbound)

Overview of TAKMI

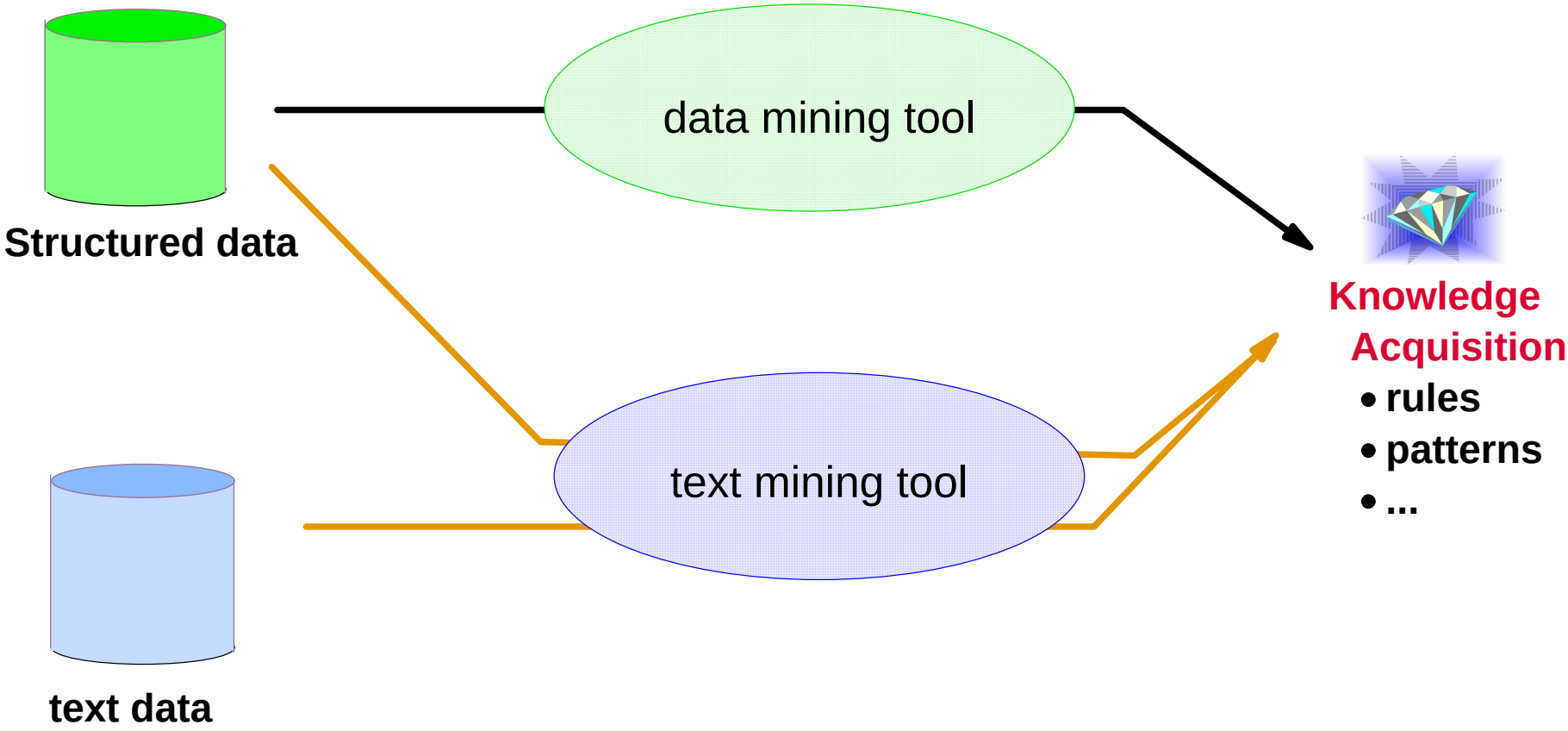
How does it work?

Basic Flow of TAKMI

After linguistic analysis, the mining engine should work interactively to offer and verify various hypotheses.

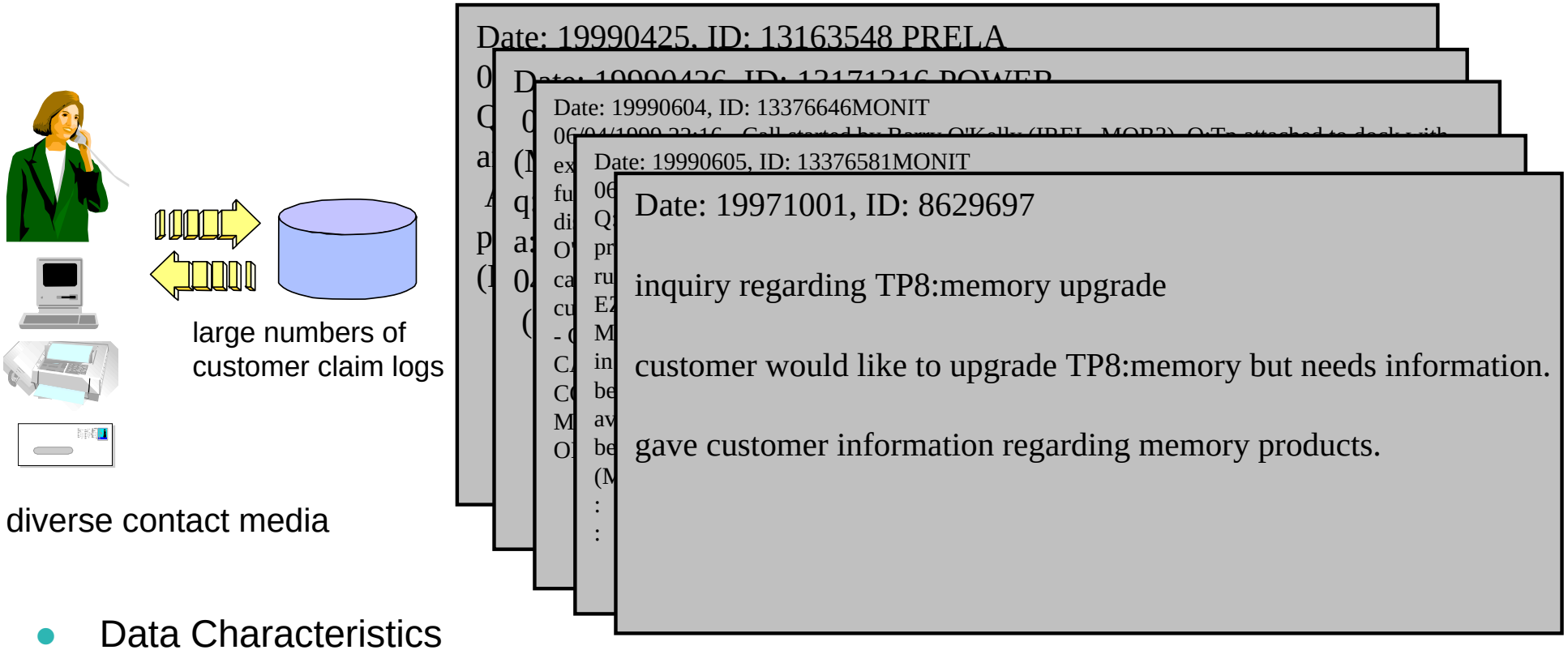


Text Mining vs. Data Mining



Application Example –Mining of Customer Contact Records–

The first application of TAKMI was to analyze customer contact records in IBM PC help centers



- Data Characteristics
  - ✓ Over 500,000 records per year.
  - ✓ Fixed fields (e.g. machine type) plus free text (body of customer comment).
  - ✓ Free text portion from 10s to 100s of words per log.
  - ✓ Overall meaning more important than case-wise analysis.

Difficulties in Analysis of Customer Contact Records

- Examples of real data
  - *A) cust wants memory upgrade info*
  - *B) install sftware...reboot of haRDdrive...make backup*
  - *C) cus said the cdrom wont open*
  - *D) cus cant click on anything*
  - *E) Told cu to C + A + D the sys*
  - *F) Got the cu to reboot the sys*
  - *G) THERE IS NOTHING WRONG WITH THE COMPUTER HE HAS A THIRD PARTY NETWORK CARD INSTALLED*

Difficulties in Analysis of Customer Contact Records

- Informal style of writing
  - ✓ Various expressions for the same concept
  - ✓ TP = T/P = ThinkPad, cu = cus = cust = customer = user
  - ✓ Ungrammatical sentences
  - ✓ Spelling mistakes
- Various types of content
  - ✓ request, question, complaint, admiration, etc.
- Various depths & strengths of concepts
  - ✓ safety issues (smoke, spark, injury, etc.)
- Multiple problems & topics in one record

Appearances of "*use*" in call records with their context of intention
(originally in Japanese)

Typical Expression	Indication of Intention				Number of Appearances
	Possible	Negation	Request	Question	
Use	N	N	N	N	1998 (56.2%)
Not possible to use	Y	Y	N	N	637 (17.9%)
Possible to use	Y	N	N	N	297 (8.4%)
Want to use	N	N	Y	N	262 (7.4%)
Is it possible to use . . . ?	Y	N	N	Y	137 (3.9%)
Do/does not use	N	Y	N	N	137 (3.9%)
Does it use . . . ?	N	N	N	Y	57 (1.6%)
Isn't it possible to use . . . ?	Y	Y	N	Y	19 (0.5%)
Others					10 (0.3%)
Total					3554 (100%)

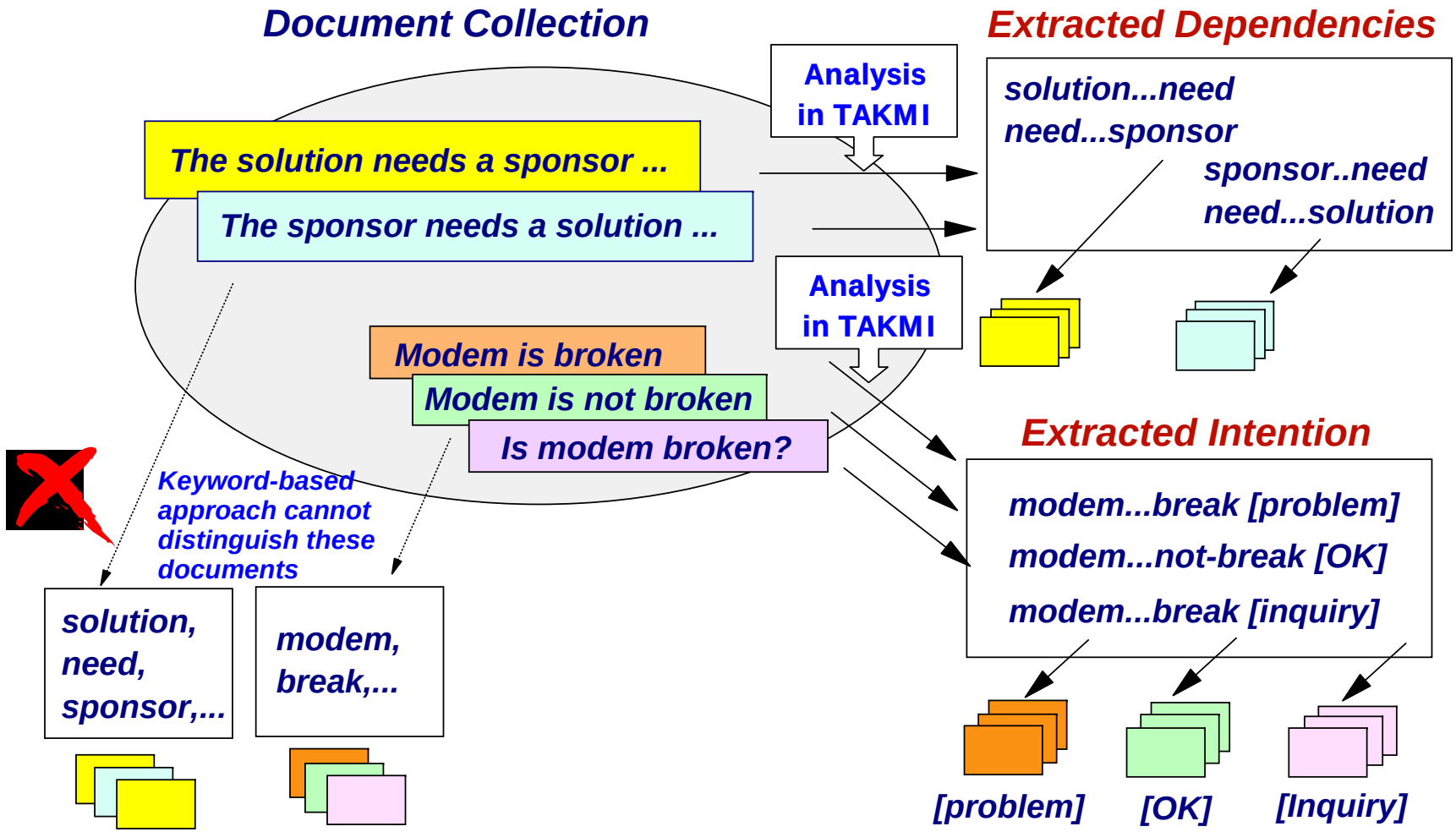
NLP for TAKMI

- Robust NLP framework for noisy data with:
  - ✓ Replacement of variants with canonical forms
    - TP = T/P = ThinkPad, cu = cus = cust = customer
  - ✓ Dependency analysis to capture sentence level information
    - subject...predicate, predicate...object
  - ✓ Intention analysis to capture types of content
    - request, question, complaint, admiration, etc.
  - ✓ Assignment of semantic categories to extracted items
    - hardware, software, problem, software...problem

Linguistic Analysis in TAKMI

Crucial problem: Text representations for statistical analysis

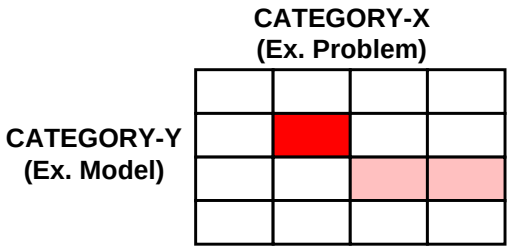
The most important issue for text mining is how to represent the textual data content in order to apply statistical analysis.



TAKMI Mining Functions

TAKMI provides various functions to capture and recognize meaningful trends

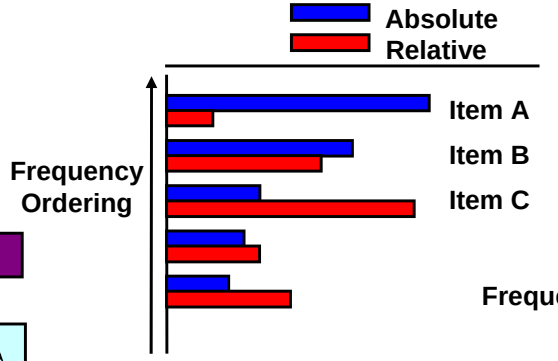
Two Dimensional Association



Easy-to-understand interface to correlation between two items (e.g. model vs.. problem)

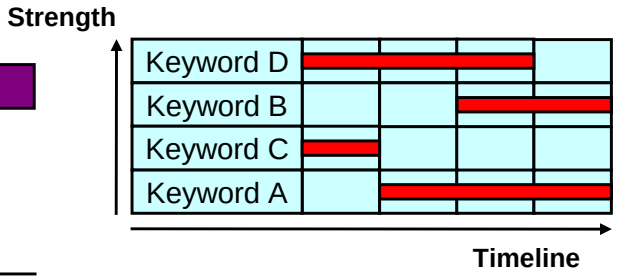
Distribution Analysis

Frequency : Number of documents including each item.



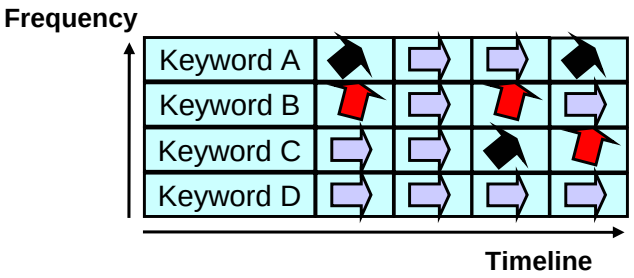
Itemized overview of the selected document collection. Relative frequency ordering facilitates collection-specific items, keywords, and so on.

Topic Extraction



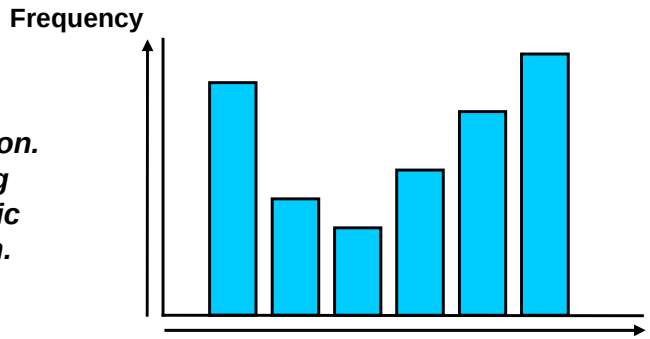
Easy-to-understand interface to salient topics over time.

Trend Analysis



Easy-to-understand interface to trend and popularity analysis

Chronological Analysis



Easy-to-understand interface to chronological distribution analysis

Application Examples of TAKMI

Application Examples of TAKMI

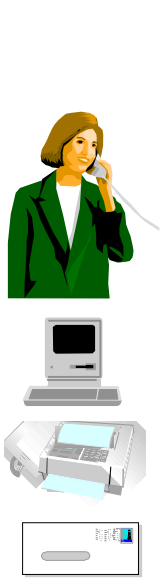
- Customer contact records in inbound call centers
  - ✓ Where customers call in for requests, questions, and complaints
  - ✓ Typically Help Centers
- Customer contact records in outbound call centers
  - ✓ Where corporate agents call customers for telemarketing
  - ✓ Can be applied to sales agent reports
- Patents
- Bioinformatics
- Others

Application Examples of TAKMI

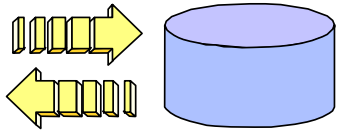
Text Mining in Inbound Call Centers

Application Example –Mining of Customer Contact Records–

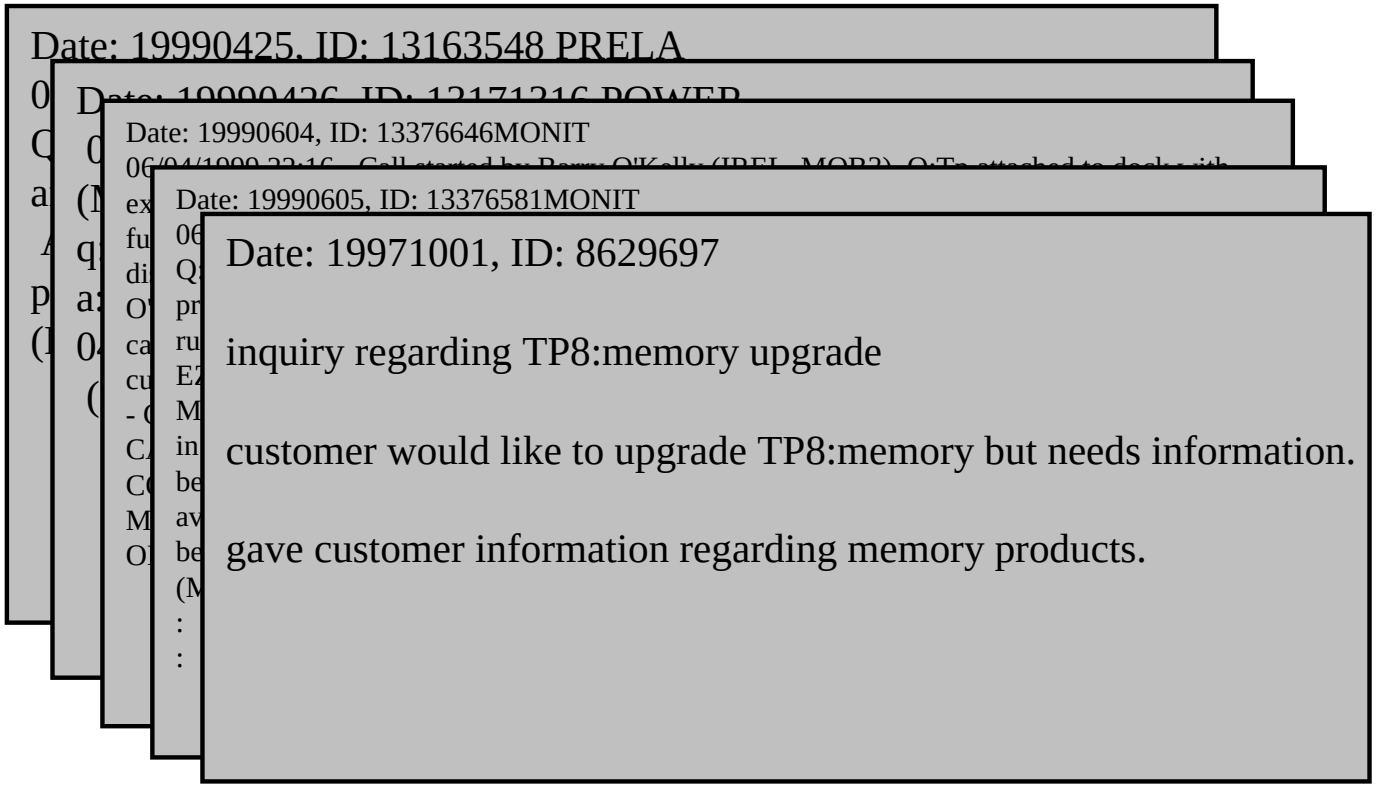
The first application of TAKMI was to analyze customer contact records in IBM PC help centers



diverse contact media



large numbers of customer claim logs



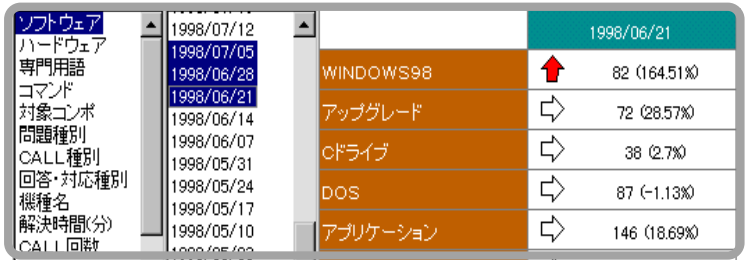
- Data Characteristics
  - ✓ Over 500,000 records per year.
  - ✓ Fixed fields (e.g. machine type) plus free text (body of customer comment).
  - ✓ Free text portion from 10s to 100s of words per log.
  - ✓ Overall meaning more important than case-wise analysis.

Mining Example –Trend Discovery and Analysis–

TAKMI is useful for detecting trends, analyzing their causes for appropriate actions, and verifying results of the actions.

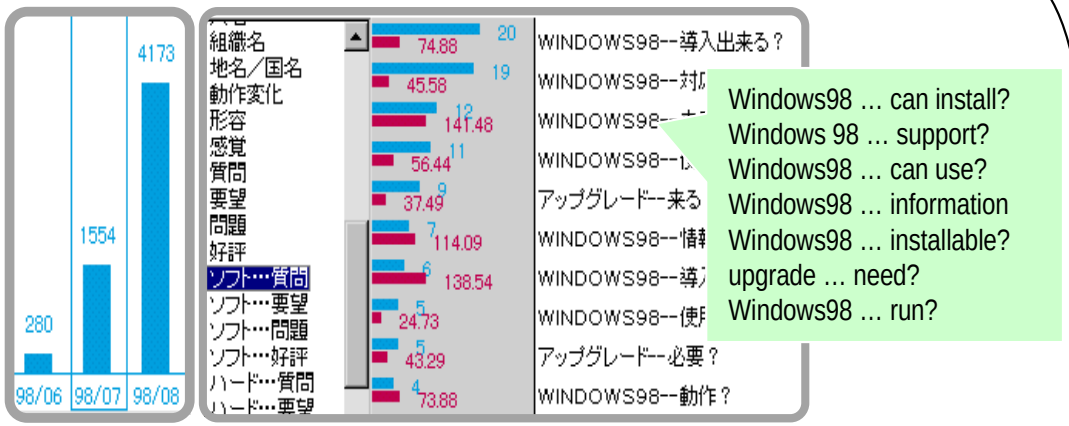
1. Find Query with Largest Increase

- *From mid-June, queries on Win98 increase sharply.*



2. Analyze Cause of Increase

- *Sudden increase of July attributed to queries on Windows 98 installation.*

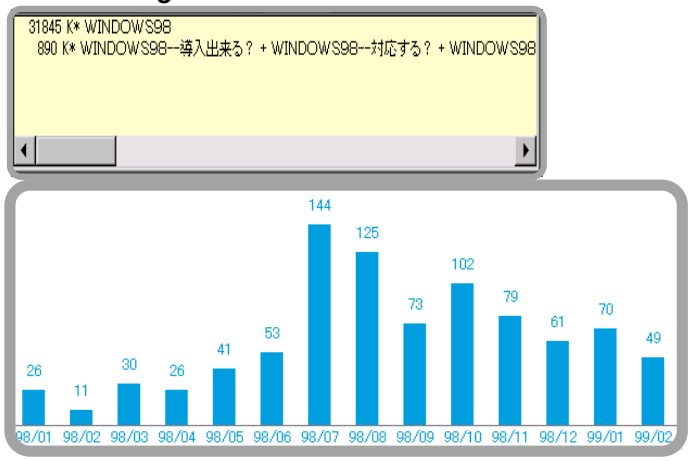


3. Action

- *Provide Win98 FAQ on installation*

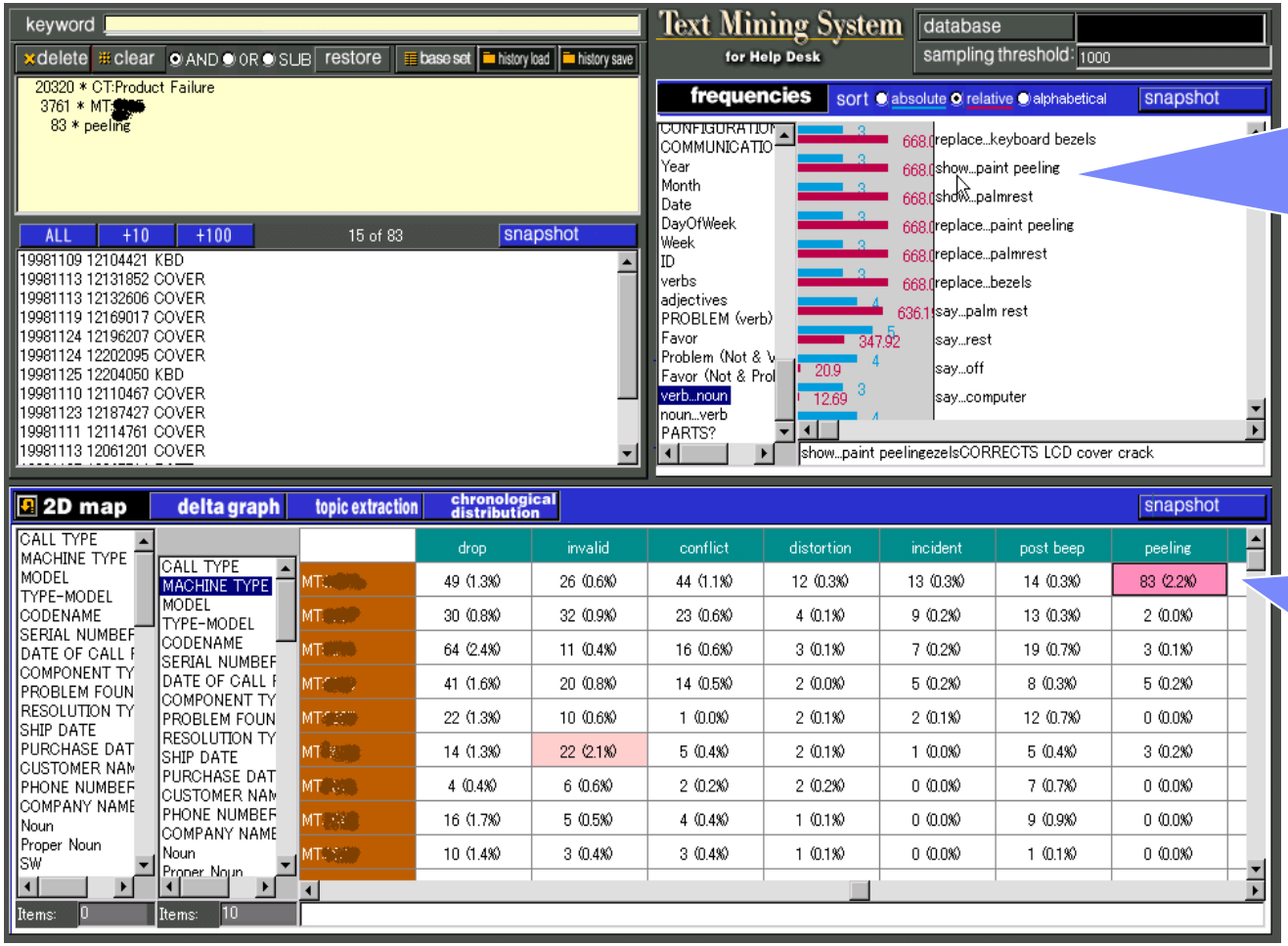
4. Evaluation

- *Queries on Windows 98 decreased after August.*



Mining Example –Problem Detection–

In PC Help Centers, TAKMI has been most effective for finding product failures in their early stages. The PC Help Center in Raleigh, NC, estimated one finding saved millions of dollars



2) Interactive mining to find features of the focused set of data

Relevant concept:
show…paint peeling
replace…palm rest

1) Automatically detected problem: "peeling" for Machine Type XXX

This problem can be focused on by clicking in this cell.

Application Examples of TAKMI

Text Mining in Outbound Call Centers

(Application for Marketing)

Analysis of Outbound Call Records for ibm.com Center in Japan

- Ibm.com Outbound call center in Chiba, Japan covers telemarketing for customers in small and medium business
- Activities are stored in a Marketing and Sales Management (MSM) system that contains
  - ✓ Customer profile
  - ✓ Contact history with brief overview of what was talked about in each contact as notes for the next contact
  - ✓ Etc.
- This project started as a request to take advantage of the huge amount of information stored in the MSM to improve the telemarketing business

Objective of Outbound Call Analysis

- **Improve productivity of marketing activities by taking advantage of data stored in MSM**
  - ✓ by mining information on
    - **Successful sales patterns**
      - How should they approach customers and propose products/services?
    - **Successful agents**
      - What makes them different from other agents?
    - **Customer information**
      - Which customers are good targets for specific campaigns?
- **Marketing activities based on data instead of intuition**

Overview of Outbound Call Analysis

- Feasibility study with approximately 210,000 records
- Obtained encouraging results on
  - ✓ Customer information
    - Which customers are good targets of this campaign?
  - ✓ Successful agents
    - What makes them different from other agents?
- Difficulty in analyzing
  - ✓ Successful sales patterns
    - How should they approach customers and propose products/services?

  Special problems in the data:
    - No link between contacts and activities for analysis of successful activities
    - Changes of time stamps due to system transition

Results of Outbound Call Analysis

- Analysis of customer information
  - ✓ on IT-related inventory
  - ✓ to analyze good target customers for a specific campaign

  The customers' use of our own products and services has been recorded, but collecting information about the customers' uses of competitor's products often clarifies the customers' real requirements and may uncover business opportunities.
- Analysis of successful agents
  - ✓ What makes them different from other agents?

Mining of customer information

Extraction of IT-related Inventory Information from Contact Records

〔Contact record with Mr. I at S Corp.〕
ご挨拶でTEL。ホストはXXX社のサーバーを使用。・・・

Call for greeting. They are *using* **XXX**'s ***server*** for host…

Company Name	Product Name	Customer Org	Status
XXX	server	S Corp.	done

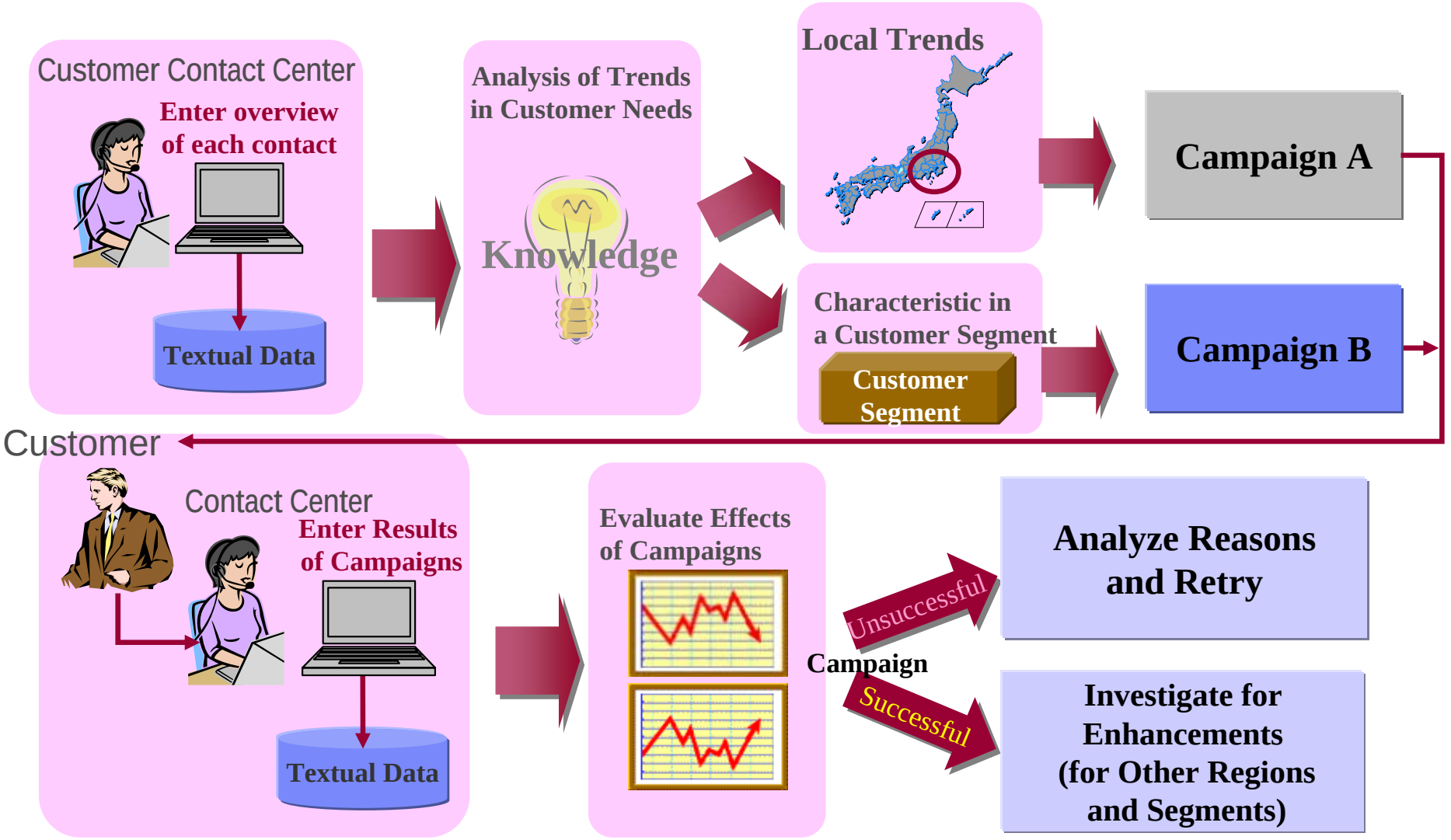
〔Contact record with Mr. H at K Corp.〕
今回の移行の件についてもXXX社の直販営業からコンサル、SVR統合の提案を受けている。

They received *proposal* from XXX's salesman on consulting and ***SVR*** integration in regard to the coming transition.

Company Name	Product Name	Customer Org	Status
XXX	server	K Corp.	potential

Application for Campaign Management

Application of TAKMI to campaign management makes it possible to analyze customers' attitudes toward each product and service



Mining information on successful agents

Comparison among successful agents

List agents by use of "*thank*" in their records

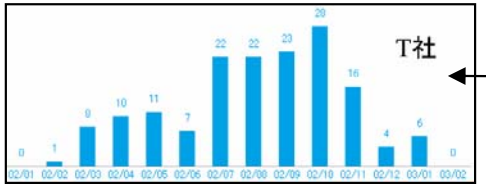
Rank	Type	Agent	No. of Records	Relative Freq.
1.		Agent:AAA	146	2.66
2.	M	Agent:ABC	130	2.31
3.	M	Agent:BBB	220	2.11
4.	M	Agent:CCC	120	2.0
5.	M	Agent:DDD	118	1.85
6.	M	Agent:EEE	148	1.84
7.		Agent:FFF	76	1.81
8.		Agent:GGG	80	1.56
9.	M	Agent:HHH	89	1.49
10.		Agent:III	18	1.49
11.		Agent:JJJ	35	1.44
12.	M	Agent:KKK	100	1.42
13.	M	Agent:LLL	60	1.17
14.	M	Agent:MMM	65	1.13
15.	M	Agent:NNN	60	1.09
16.		Agent:OOO	27	1.02
17.	M	Agent:PPP	36	1.01
18.	M	Agent:QQQ	43	0.0

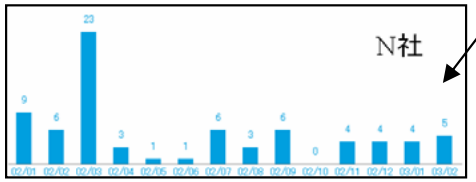
Rank	Type	Agent	No. of Records	Relative Freq.
19.		Agent:RRR	38	0.0
20.		Agent:SSS	37	0.0
21.		Agent:TTT	56	0.0
22.		Agent:UUU	54	0.0
23.		Agent:VVV	34	0.0
24.		Agent:WWW	30	0.0
25.		Agent:XXX	51	0.0
26.		Agent:YYY	19	0.0
27.	M	Agent:ZZZ	47	0.0
28.	M	Agent:aaa	12	0.0
29.	M	Agent:bbb	11	0.0
30.		Agent:ccc	10	0.0
31.		Agent:ddd	9	0.0
32.		Agent:eee	8	0.0
33.	M	Agent:fff	8	0.0
34.	M	Agent:ggg	3	0.0
35.	M	Agent:hhh	2	0.0

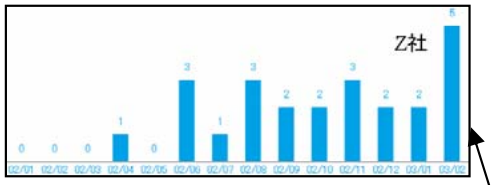
Mining information on successful agents

Comparisons among successful agents

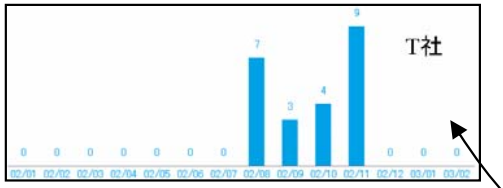
Continuous contact patterns

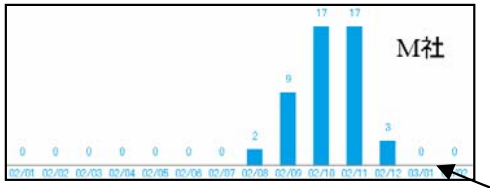






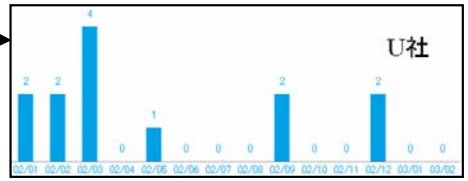
Concentrated contact patterns

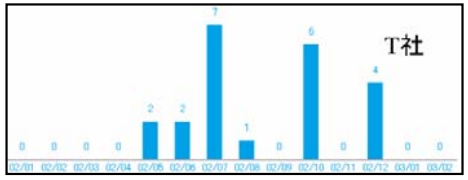


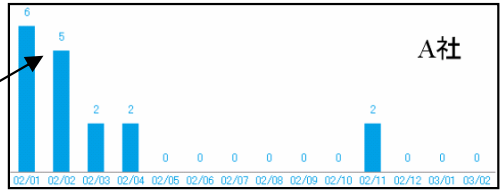


List of agents sorted by use of "*thank*" in their records

AAA	146(2.66)	RRR	38(0.0)
M:ABC	130(2.31)	SSS	37(0.0)
M:BBB	220(2.11)	TTT	56(0.0)
M:CCC	120(2.0)	UUU	54(0.0)
M:DDD	118(1.85)	VVV	34(0.0)
M:EEE	148(1.84)	WWW	30(0.0)
FFF	76(1.81)	XXX	51(0.0)
GGG	80(1.56)	YYY	19(0.0)
M:HHH	89(1.49)	M:ZZZ	47(0.0)
III	18(1.49)	M:aaa	12(0.0)
JJJ	35(1.44)	M:bbb	11(0.0)
M:KKK	100(1.44)	ccc	10(0.0)
M:LLL	60(1.17)	ddd	9(0.0)
M:MMM	65(1.13)	eee	8(0.0)
M:NNN	60(1.09)	M:fff	8(0.0)
OOO	27(1.02)	M:ggg	3(0.0)
M:PPP	36(1.01)	M:hhh	2(0.0)
M:QQQ	43(0.0)		







Successful agents are indicated by "M:"

Mining information on successful agents
Comparisons among successful agents

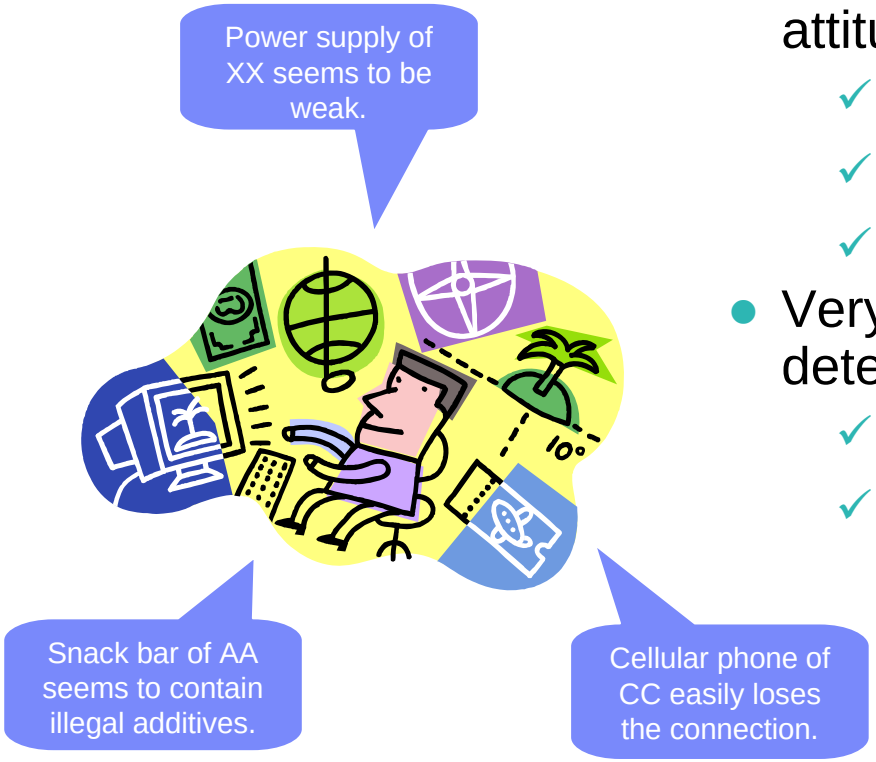
- As a result of analyzing use of "*thank*" in contact records, we found successful agents can be classified into two groups
  - ✓ Concentrated contact pattern group
  - ✓ Continuous contact pattern group
- Skill levels of successful agents in the concentrated contact pattern group are very high
- Agents in the continuous contact pattern group can be successful without high skills

➡ Use this insight for agent education

Sentiment Analysis

Sentiment Analysis
Application Overview

User's Reaction in Internet Era



- Reputation messages appear on many places in Internet, which is not an official channel to companies.
  - ✓ Bulletin Boards, and Homepages
- These information affects potential user's buying attitude.
  - ✓ AA PC is likely to be broken in a few years.
  - ✓ Power cell of BB cellular phone is not functioning.
  - ✓ A child suffered food poisoning at CC restaurant.
- Very Important for corporate activities to quickly detect these information.
  - ✓ Correction of Wrong Information
  - ✓ Improvement of Products and Services

User Survey

- User survey is usually used when new products and services are planned. Fixed-style questionnaire is hard to get free opinion, since questions are fixed based on assumptions.
- Free-style questionnaire is ideal, but is hard to read all opinions by human.
  - ✓ Need a tool to assist this activity.

Fixed-Style Questionnaire

Q1. Have you ever used a product called XXX.
○ Yes
○ No

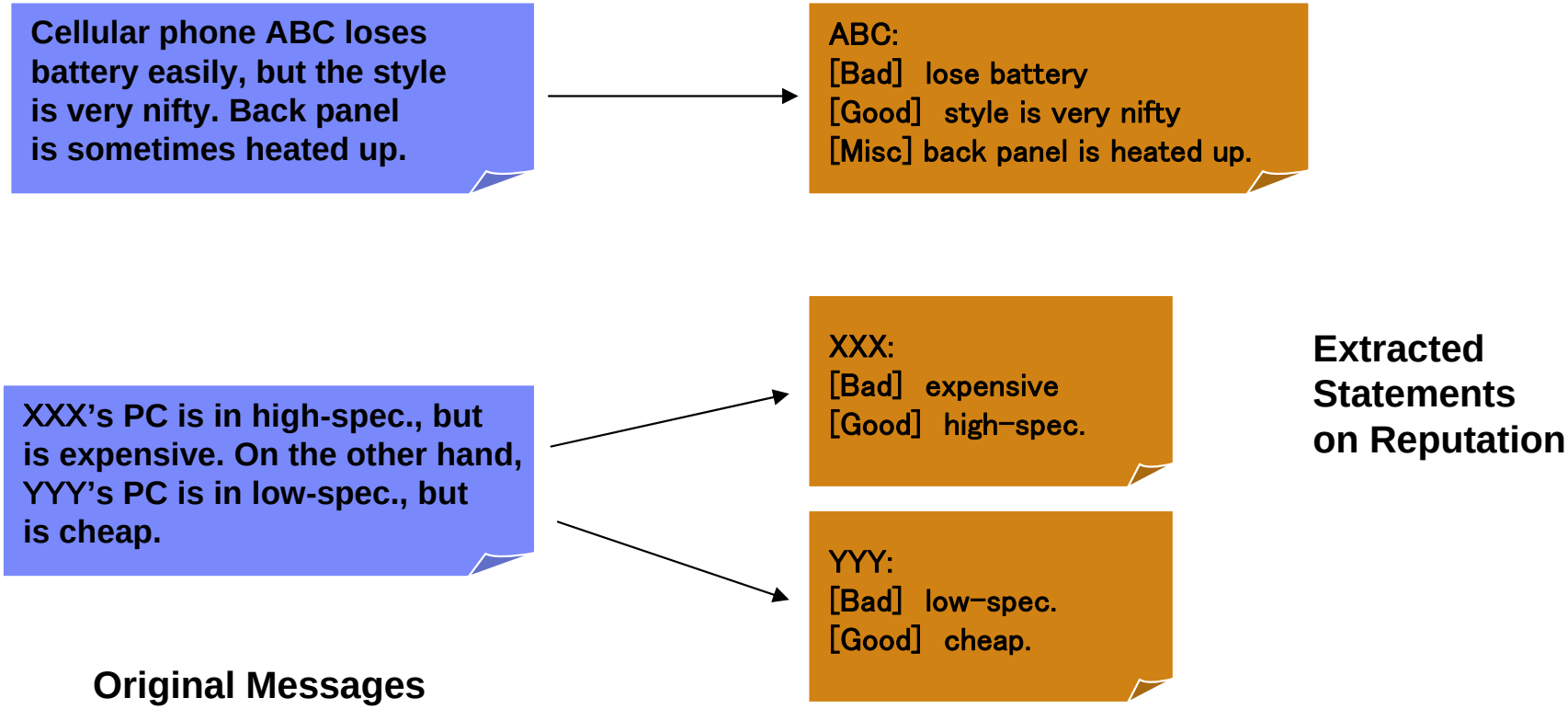
Q2. How do you feel when you use it.
A. Very easy
B. Hard to use
C. Others
…..

Free-Style Questionnaire

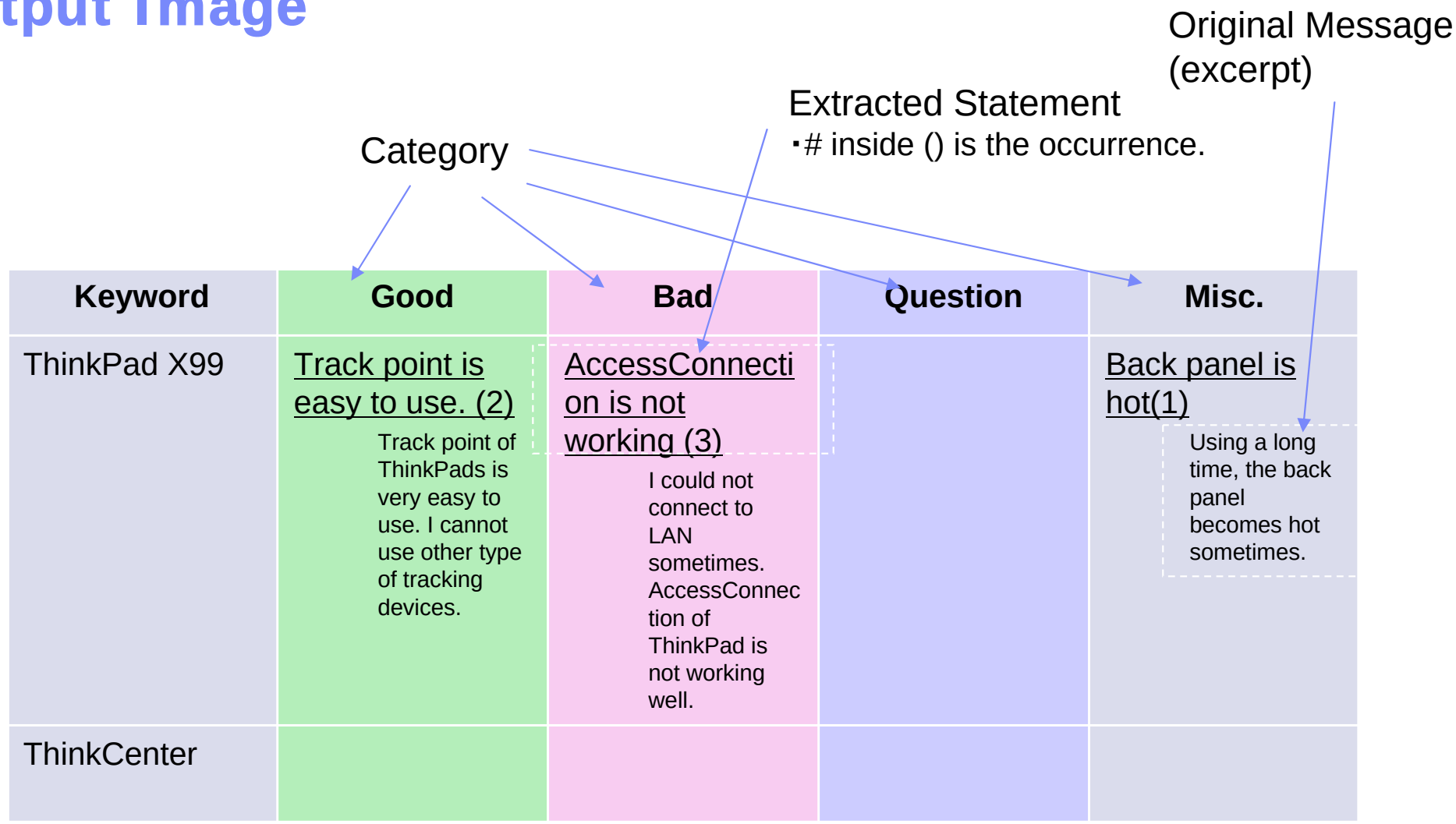
Q. Write what you think and feel when you use the product called XXX.
（E.g.：At first, it is slightly hard to learn to use XXX. But, now it is very easy to use, and it is an essential part of my life.
…..

TAKMI Reputation Mining

- Based on TAKMI's intention analysis function, this extracts key statements related to reputation.
- Key statements are assigned a category such as positive or negative.



Output Image



Keyword	Good	Bad	Question	Misc.
ThinkPad X99	Track point is easy to use. (2) Track point of ThinkPads is very easy to use. I cannot use other type of tracking devices.	AccessConnection is not working (3) I could not connect to LAN sometimes. AccessConnection of ThinkPad is not working well.		Back panel is hot(1) Using a long time, the back panel becomes hot sometimes.
ThinkCenter				

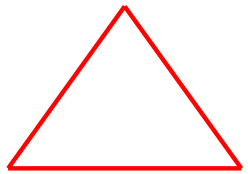
Note: The above table just shows the image of the output of this system. It does not assure that these statements are exactly extracted.

Sentiment Analysis
Technical Overview

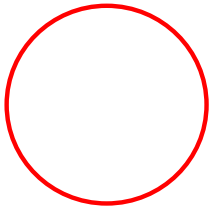
Issues of Sentiment Analysis (1)

- Document level analysis of favorability is difficult
  - Interpretation of opinions can be debatable even for humans
  - Documents often contains both favorable and unfavorable comments
- In practice, instances of favorability are often more useful than simple polarities of favorable vs. unfavorable

product A = +10 (very favorable)
product B = - 2 (slightly unfavorable)



product A = tough, lightweight, slim
product B = heavy, thick



Issues of Sentiment Analysis (2)

- Analysis of semantic relationships between subject terms and sentiment terms is required to assign favorability properly.

  XXX wins over YYY.

  - ✓ Favorable for *XXX*
  - ✓ Unfavorable for *YYY*
- Definitions of sentiment terms should be more than the polarity of favorable/unfavorable.

Basic Framework of our Approach

- Detect positive/negative mentions of the subject within *local contexts*.
- Handle multiple regions within a document, each with a favorable or unfavorable expression.
- Determines *polarity* of sentiment.
- Uses a *sentiment in skeleton* format for easy review by users.

ThinkPad is very expensive, but I love IBM products.

-1 ThinkPad (ThinkPad)---be (is very)---expensive (expensive)

1 love (love)---IBM (IBM products)

Definition of Sentiment Expressions

Sentiment Expressions:

- Adjective
  - <favorable> ***good, helpful, capable, super, …***
  - <unfavorable> ***bad, helpless, inaccurate, crude, …***
- Adverb
  - <favorable> ***well***
  - <unfavorable> ***badly, poorly, …***
- Noun
  - <favorable> ***progress, benefit, …***
  - <unfavorable> ***fault, injury, blame, …***

Definition of Sentiment Expressions

- Enrichment of definition for sentiment verbs
  - ✓ Sentiment verbs
    - directly indicate (un)favorability toward their argument
  - ✓ Transition verb:
    - does not determine sentiment
    - requires sentiments in its argument phrases
  - ✓ An argument can be subject, object, complement, or PP associated with the verb

gVB lead sub // *Product XYZ leads the segment.*
bVB fine obj // *Regulators fined Company_S $5M for misleading stock research.*
tVB provide obj sub // *Company_I provides a good working environment.*
tVB prevent obj ~sub // *XXX prevents troubles.*

Definition of Sentiment Expressions

- Introduction of neutral phrase to ignore favorability in idiomatic phrases
  - ✓ ***crude*** *oil*, *jet* ***lag***, *with* ***respect*** *to*, *etc.*

bJJ crude
nNN crude oil
bJJ lag
nNN jet lag

Definition of Sentiment Expressions

POS	Total	Positive	Negative	Neutral
Adjective	2,465	969	1,495	1
Adverb	6	1	4	1
Noun	576	179	388	9
Sentiment verb	357	103	252	2
Transfer verb	109			

Basic Framework of our Approach

Use of Part-of-speech (POS) tagging

- POS tagging enables
  - ✓identifying sentiment terms properly
    - well:adverb = favorable
      XXX works well.
    - well:noun ≠ favorable
      Well of Wisdom
    - like:verb = favorable
      I like YYY.
    - like:preposition ≠ favorable

Basic Framework of our Approach
Use of shallow parsing

- Shallow parsing enables
  - ✓ identifying phrase boundaries
    - Noun phrase, Verb phrase, etc.
  - ✓ identifying local dependencies between phrases
    - Subject of verb phrase
    - Object of verb phrase

  (based on Talent System [Neff et al. 2003])

Basic Framework of our Approach

Example of shallow parsing output

Input sentence:

This sentence might be good for representing an example of shallow parsing output.

Output result:

(Subject (NP ***This*** [this|DT] ***sentence*** [sentence|NN]))
(VG ***might*** [may|MD] ***be*** [be|VB])
(AdjP ***good*** [good|JJ] ***for*** [for|IN])
(VG ***representing*** [represent|VBG])
(NPP
  (NP ***an*** [an|DT] ***example*** [example|NN])
  (PP ***of*** [of|IN]
    (NP ***shallow*** [shallow|JJ] ***parsing*** [parsing|NN] ***output***[output|NN])))

Results of our sentiment analysis on test corpus from Web pages

Dataset 1: Auto, IT, Music, Oil, and Camera industries

Cases: 175 cases (118 positives, 57 negatives)

Sentiment detected: **53** cases

Number of correct cases = **50**

Precision = (50/53) = 94.3%

Recall = (50/175) = 28.6%

$$\text{Precision} = \frac{\text{Number of correct cases}}{\text{Number of } \textbf{system} \text{ assigned sentiments}}$$

$$\text{Recall} = \frac{\text{Number of correct cases}}{\text{Number of } \textbf{human} \text{ assigned sentiments}}$$

Results with Open Test Corpus

Dataset 2: 2,000 Camera Reviews

About half of them are neutral

Sentiment detected: 255 cases

Number of correct cases = **241**

Precision = (241/255) = 94.5%

Recall ≒ (241/1000) ≒ 24.1%

Dataset 3: 476,126 Web pages

Focused on 1,198 pages that mention a product

3,804 subject references (names of ten medicines)

Sentiment detected: 103 cases

Precision = 91%

Applications of Sentiment Analysis

What can be done with sentiment analysis with less than 50% recall?

- Disadvantages of Sentiment Analyzer
  - ✓ Misses over half of the sentiments
  - ✓ Cannot compete with human analysts in depth and accuracy of judgments for each location
- Advantages of Sentiment Analyzer
  - ✓ Detect polarity (positive/negative) of each mention of a sentiment
  - ✓ Extract skeletons to look over the results
    - Allows focusing on critical terms such as "*hate*", "*violate*", etc.
  - ✓ Automatically run on large numbers of documents
    - Possible to go through billions of pages or locations

Applications of Sentiment Analysis with high precision and low recall

- Capturing trends on sentiments (similar to analysis by sampling)
  - ✓ Compare sentiments among corporations/products
  - ✓ Detect bad rumors before they spread
- Finding important documents to be monitored
  - ✓ Documents worth tracking often contain many sentiment expressions

  → chances of finding such pages should be relatively high even with the low recall.

Capturing Trends on Sentiments

Does output of sentiment reflects trends?

Comparison of numbers of sentiments on camera brands detected by humans and by this system

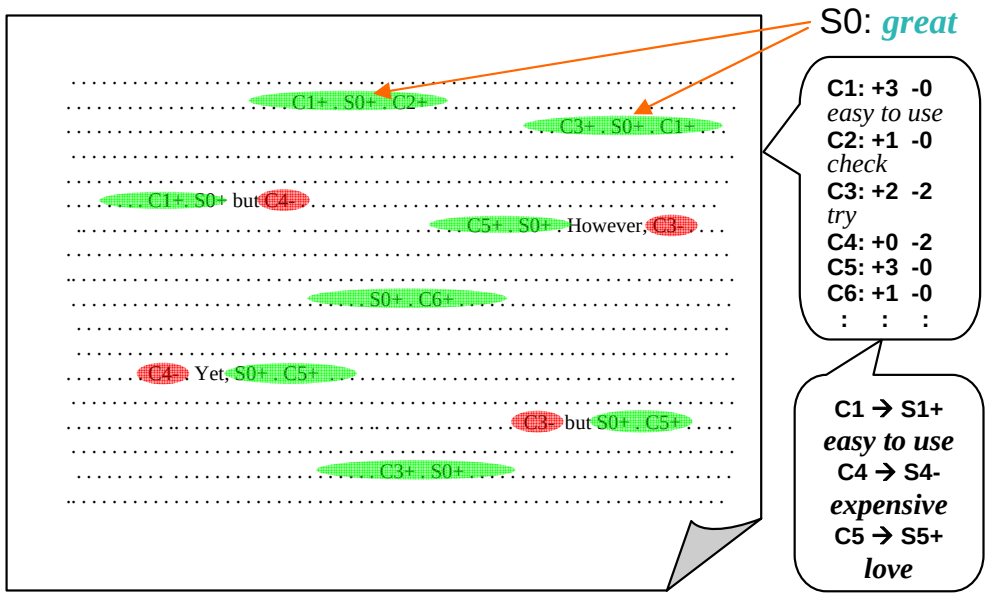
	polarity	**brand *A***	**brand *B***	**brand *C***	**brand *D***
Human	favor.	437	169	80	39
	unfav.	70	65	51	41
System	favor.	52	22	9	3
	unfav.	4	5	2	1

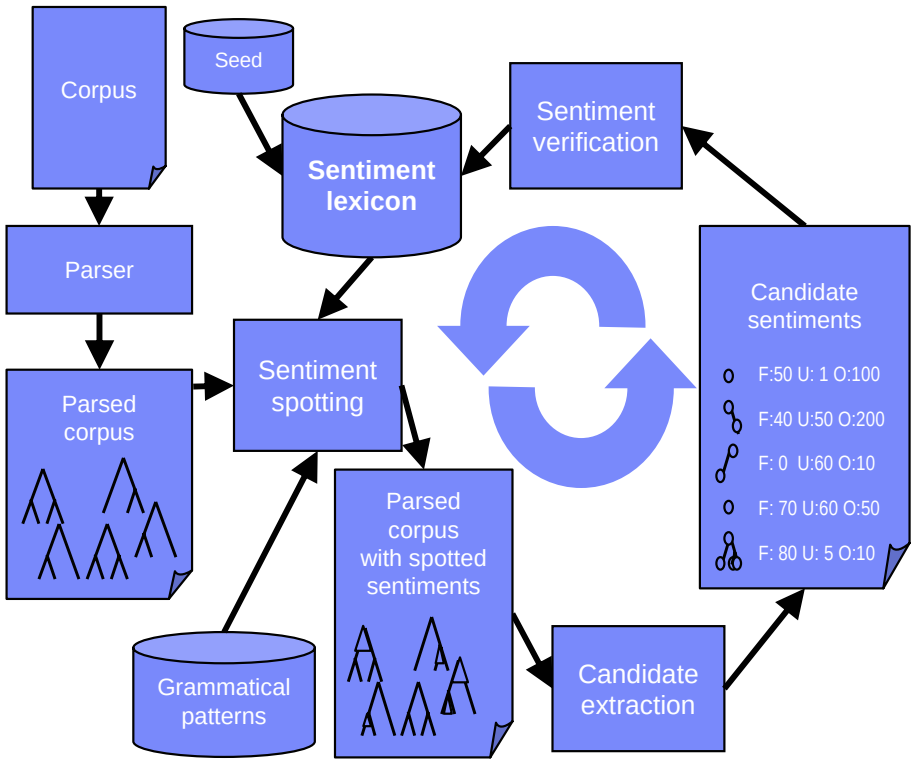
Sentiment Analysis
Progress

ESPER (Extraction of Sentiment and Preference Expressions)

- **Technology to identify what is considered favorable or unfavorable**
- ➔ Unsupervised acquisition of sentiments from domain corpora by taking advantage of the following characteristics of sentiment expressions
  - ✓ Consecutive sentiment expressions tend to have the same polarity of favorability, unless a clue term (such as an adversative conjunction) signals a polarity switch

Example: *I have been very **impressed** with this camera. It takes **great** pictures and is very **easy** to use. But the price is too **expensive**.*





ESPER Applications

- **Automatic acquisition of domain-dependent expressions**
  used for sentiment analysis and reputation mining

Domain	Polarity	Expressions automatically extracted from postings in bulletin board data
Digital camera (18,000 postings)	favorable	*beautiful, helpful, easy to hold, easy to use, etc.*
	unfavorable	*disturb, poor image quality, create noise, more noise, etc.*
Movie (75,000 postings)	favorable	*draw smiles, neat, feel like crying, terrifying, stay in memory, fearful, etc.*
	unfavorable	*got bored, bad, tedious, predictable, difficult, stop crying, etc.*

- **Identification of changes in people's preferences over time**
  by identifying what is considered to be positive and negative among various expressions flexibly in unannotated textual data

	Preference expressions	Appearances in 2002 data (25,659 postings)	Appearances in 2004 data (35,669 postings)
Increasing in positive contexts	*LCD panel is large*	0	12 (12 positive)
Increasing in negative contexts	*no tripod mount*	0	4 (4 negative)
Decreasing in positive contexts	*size is small*	5 (3 positive, 2 negative)	1 (1 positive)
Decreasing in negative contexts	*slow to start*	5 (1 positive, 4 negative)	0

Conversation Mining

Overview

- Initial Application of ASR (Automatic Speech Recognition) Transcribed Records of a Call Center to TAKMI
- Current Status
  - ✓ Over 110K Records
  - ✓ No Customization in Lexicon
  - ✓ No Distinction between Customer and Agent
- Current Findings
  - ✓ Difference between Written Language and Spoken Language
  - ✓ Difference in Noise

Summary

Summary

- TAKMI
  - ✓ Overview
  - ✓ Application Examples
    - Customer Contact Records
- Sentiment Analysis
  - ✓ Application Overview
    - Reputation Mining
  - ✓ Technical Overview
  - ✓ Progress
- Conversation Mining
  - ✓ Overview
  - ✓ Demo

References

References

- TAKMI
  - ✓ Text analysis and knowledge mining system, T. Nasukawa T. Nagano http://www.research.ibm.com/journal/sj/404/nasukawa.html
- MedTAKMI
  - ✓ A text-mining system for knowledge discovery from biomedical documents N. Uramoto, H. Matsuzawa, T. Nagano, A. Murakami, H. Takeuchi, and K. Takeda http://researchweb.watson.ibm.com/journal/sj/433/uramoto.html
- Term aggregation
  - ✓ Murakami, A. and Nasukawa, T. "Term aggregation: mining synonymous expressions using personal stylistic variations", International Conference on Computational Linguistics (COLING), 2004.
- Sentiment analysis
  - ✓ Nasukawa, T. and Yi, J. "Sentiment analysis: capturing favorability using natural language processing", International Conference on Knowledge Capture, pp.70-77, 2003. (http://portal.acm.org/citation.cfm?id=945645.945658)
- Deeper sentiment analysis
  - ✓ Kanayama, H., Nasukawa, T. and Watanabe, H. "Deeper sentiment analysis using machine translation technology", International Conference on Computational Linguistics (COLING), pp.494-500, 2004.
- Text Mining project in IBM Tokyo Research Laboratory
  - ✓ http://www.research.ibm.com/trl/projects/textmining/index_e.htm