

# Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction



Fredrik Espinoza, Ola Hamfors, Jussi Karlgren, Fredrik Olsson, Per Persson, Lars Hamberg, Magnus Sahlgren

## TARGET TASK

Given a set of utterances such as consumer reviews, customer feedback, open answers to questionnaires, cluster them rapidly and interactively into domain-dependent themes of interest to provide support for further decision making and action.

Open questions and textual feedback allows an analyst to explore what is not predictable; the ambiguity, vagueness, and fluidity over time of human language use allows respondents and correspondents to express what might not be expected by the analyst.

This sort of data requires qualitative analysis. Combining qualitative analysis with quantitative analysis requires the text items to be systematically processed by a human analyst. The system presented here is a support system for such analysis.

## OPEN ANSWERS

### Hotel reviews: topical themes

- (1)
  - a. I would definitely recommend this hotel, the location was great!
  - b. Had I known, I would NOT have chosen this hotel for my busy work visit in which I needed quiet time in hotel to do work.
  - c. Modern, stylish hotel with numerous, pretty decent restaurants in the area!

### "What makes text trustworthy?": unexpected themes

- (2)
  - a. The appearance of the text, the quality of its design and polish.
  - b. How enjoyable and fun it is, how it addresses its readers, and who has written it.
  - c. Who wrote it and why.
  - d. Does it speak to me?

### Hotel reviews: synonyms

- (3)
  - a. The staff were very friendly and helpful.
  - b. The staff was courteous and professional, and they gave the impression that hospitality was something they enjoyed expressing.
  - c. The staff was personable and demonstrated a true thankfulness for your business.
  - d. The breakfast was always fine and we enjoyed a light breakfast every morning of a bowl of fruit together with a choice of a bagel, toast or croissant.
  - e. However the hotel did offer free pastries, muffins, fruit, coffee, and juices every morning.
  - f. There was no restaurant when we were there but they did offer coffee and pastry in the AM.

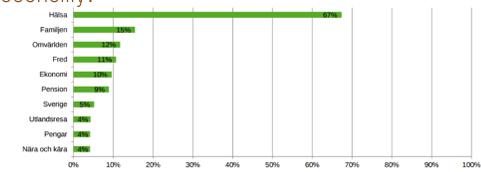
## What makes airline passengers happy?

20 000 consumer reviews of 22 airlines, clustered and processed with tonality analysis, show how airline passengers wish to be seen and met as individuals: the happiest passengers complained mostly about meals; the unhappiest about service. Satisfaction with staff service was a driver for satisfaction with other aspects, such as seat comfort of the seat, meals, and overall experience.

Passenger Satisfaction (Descending %)	Score (%)	Areas for Improvement (Descending %)	Service		
			Happy (%)	Unhappy (%)	Net Happy (%)
China Southern Airlines	0.62	Food (3.8), Service (3.1), Seat (2.6)	70	3.1	66.9
All Nippon Airways	0.55	Food (11.2), Seat (7.6), Service (6.9)	48.2	4.9	43.3
Lufthansa	0.43	Food (8.0), Seat (7.5), Entertainment	45.2	3	42.2
Qatar	0.422	Food (11.0), Seat (9.2), Service (4.5)	28	4.5	23.5
Singapore Airlines	0.395	Food (13.0), Seat (10), Entertainment	48.6	4.8	43.7
Norwegian					
Cathay Pacific Airways					
Thai Airways					
Qantas					
Emirates					
Turkish Airlines					
Virgin Atlantic					
Air France					
Southwest Airlines					
SAS					
Delta Airlines					
Ethiopia					
China Eastern Airlines	0.063	Service (24.3), Food (20.7), Seat (10.1)	18	24.3	-6.3
British Airways	-0.013	Service (19.2), Seat (18.8), Food (17.6)	27	19.2	7.8
Air China	-0.05	Service (23.4), Food (20.0), Seat (9.7)	19.6	29.4	-9.8
United Airlines	-0.132	Service (15.3), Food (10.0), Seat (9.3)	16	15.2	0.8
American Airlines	-0.275	Service (27.3), Food (8.4)	14.7	27.3	-12.6

## "What do you most wish for the coming year?"

In a survey of senior citizens with 14,793 responses clustered by theme and by worry and concern found major themes for health, world peace, and economy.



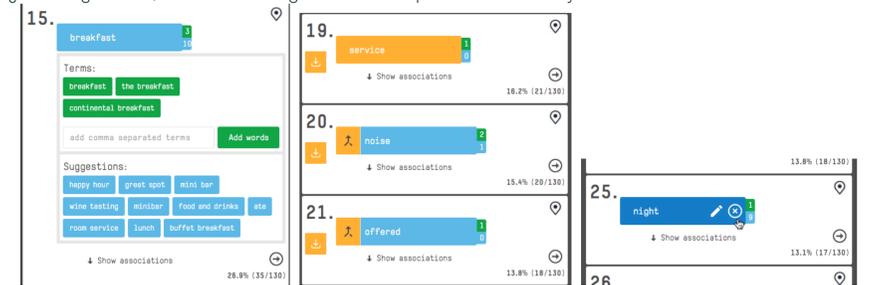
## INTERACTIVITY, NOT AUTOMATION

Our design principles are based on human language being useful as is, and on automating drudgery, not creation of insights.

- Design principle 1: Empowering analysts, not replacing them
- Design principle 2: Incremental refinement in clustering pipeline
- Design principle 3: Errors do not matter
- Design principle 4: Representation in surface terms
- Design principle 5: No dependence on outside resources

## WORDS and UTTERANCE CLUSTERS

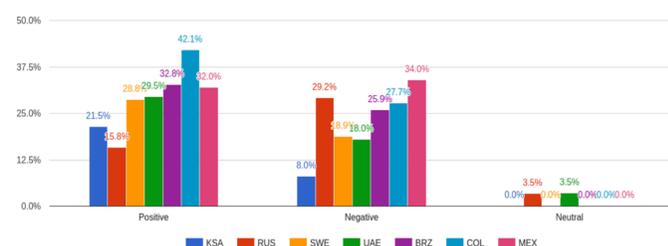
The system clusters text items by lexical statistics to create topically potentially coherent clusters for approval by the analyst. Separating them, joining them, or refining them is up to the analyst.



## Attitudes to feminism in different cultural areas

A large study on feminism with 9,800 respondents in four cultural areas: the Middle East, Latin America, Russia, and Sweden found that feminism was associated with negative gender behavioural patterns in Latin American countries and in Russia, whereas it was accepted positively as a label for progressive policies in Middle Eastern countries.

- (4) If a man or woman describes themselves as feminist, what would you think of that person? What kind of associations do you get? Is feminism positive or negative in your view? How would you describe feminism?
  - a. "Feminism is a positive concept, as women previously were discriminated against (earlier the world was sexist) whereas now women also find positions in areas which earlier were considered to be only for men."
  - b. "Feminism is neutral until it has acquired a mass character."
  - c. "I have a neutral view on this topic as each individual has their own perspective, as for me feminism shouldn't exist in today's world and education system."
  - d. "I consider feminism to be negative that it is the opposite to machismo or am I wrong?"



## CONCLUSIONS and ADVANTAGES

- scale:** full coverage of all text items
- speed:** an analyst can process more text items in a given period
- consistency:** no need to use several analysts; coding schemes can be reused; no fatigue effects
- granularity:** larger amounts of data allow for more granular analyses
- effort:** human analyst relieved from mechanical task, effort saved for intellectual task