Risk potential of words based on 20Q-games

Laszlo Pitlik (sen), Laszlo Pitlik (jun), Matyas Pitlik, Marcell Pitlik (MY-X team)

Abstract: This paper shows a calculation scheme for deriving risk potential of arbitrary human abstractions/words (like horse and/or knowledge) being guessed by a lot of human beings in order to demonstrate that the level of abstractions and therefore the level of risk potentials caused through misunderstandings can be measured based on the subjective/intuitive answers of human being to question of an online 20Q-robot. The highlighted examples could be characterized with a set of variables in a parallel way where each variable is capable of evaluating risks. Therefore, an aggregated risk potential of each word should be derived in frame of an antidiscriminative model where the objects are the words, the attributes are the description variables and the cells are the descriptive values as such. The similarity-based modelling needs hardly mathematical competences, yet it is able to estimate the appropriate risk potential for each human abstraction. The sentences contain words, so the risk potential of sentences can also be derived. The speakers having (constantly or increasingly) high-level risks through the used words in their sentences should not have the possibility of the freedom of speech?!

Keywords: measuring of abstraction levels, freedom of speech, measuring of misunderstanding potentials of human sentences

Introduction

This paper is the newest part of the series about experiences of the QuILT-based education processes. Previous articles and their annexes can be downloaded here:

1. https://miau.my-x.hu/miau/quilt/Definitions_of_knowledge.docx + annexes like:
    - https://miau.my-x.hu/miau/quilt/demo_questions_to_important_messages.docx
    - https://miau.my-x.hu/mediawiki/index.php/QuILT-IK045-Diary
    - https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK045-Diary
    - https://miau.my-x.hu/mediawiki/index.php/QuILT-IK059-Diary
    - https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK059-Diary
2. https://miau.my-x.hu/miau/quilt/reality_driven_education.docx + annexes like:
    - https://miau.my-x.hu/miau/quilt/chained-translations-legal-slang.docx
    - https://miau.my-x.hu/miau/quilt/demo_chained_translations.docx
    - https://miau.my-x.hu/miau/quilt/demos_chained_translations.docx
    - https://miau.my-x.hu/miau/quilt/forum_details.docx
    - https://miau.my-x.hu/mediawiki/index.php/QuILT-IK057-Diary
    - https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK057-Diary
3. https://miau.my-x.hu/miau/quilt/Exercises_for_critical_thinking_and_doing.docx
4. https://miau.my-x.hu/miau/quilt/st1_all.docx
5. https://miau.my-x.hu/miau/quilt/20Q.docx
6. https://miau.my-x.hu/miau/quilt/GDP_final_en.doc
7. https://miau.my-x.hu/miau/quilt/st2_all.docx
8. https://miau.my-x.hu/miau/quilt/harmony.docx
9. https://miau.my-x.hu/miau/quilt/safety-index.docx
10. (https://miau.my-x.hu/miau/quilt/20q_based_fingerprints_of_words.docx)

Parallel, there are a lot of spreadsheets supporting the needs for details: https://miau.my-x.hu/miau/quilt/?C=M;O=D

The force fields of the **magic of words** are used ever since for communication between/among human beings. The magic of words can be interpreted from different points of views: e.g. the magic of words is responsible for the misunderstandings of the text-based (verbal and/or written) communication. But parallel, the magic of words makes possible to derive new innovative ideas even based on the potential misunderstandings. The principle of KNUTH (c.f. knowledge is what can be transformed into source code – each other human activity is a kind of artistic performance) expects that we are capable of measuring the misunderstanding potential of the human words (abstractions). This paper presents a simple way to achieve the expectations of KNUTH.

20Q-games

Based on following preparation works like:

- https://miau.my-x.hu/miau/quilt/st1_all.docx
- https://miau.my-x.hu/miau/quilt/20Q.docx
- https://miau.my-x.hu/mediawiki/index.php/QuILT-20Q-raw-games
- https://miau.my-x.hu/mediawiki/index.php/QuILT-20Q-9
- **https://miau.my-x.hu/miau/quilt/20q_horse_demo.xlsx**

and based on the online services of http://20q.net/, the collection and analyses of appropriate data assets could be executed in co-operation with Students. This paper will present each relevant step of the reproducibility based on the above-highlighted XLSX-objetc.

At first, it seems to be relevant to summarize, what kind of input and output will be realized if somebody tries to play a game with the robot of 20q.net?

- it is relevant to select a thinking/cultural frame
- it is also relevant to define basic data about the human player
- the human player has to choose a word (e.g. horse or knowledge)
- and the human player has to answer the questions of the robot
- where the answers can not be arbitrary words just one from the prepared list
    - the first question is always the same and it makes possible to classify the keyword to the main categories like animal, plant, etc.

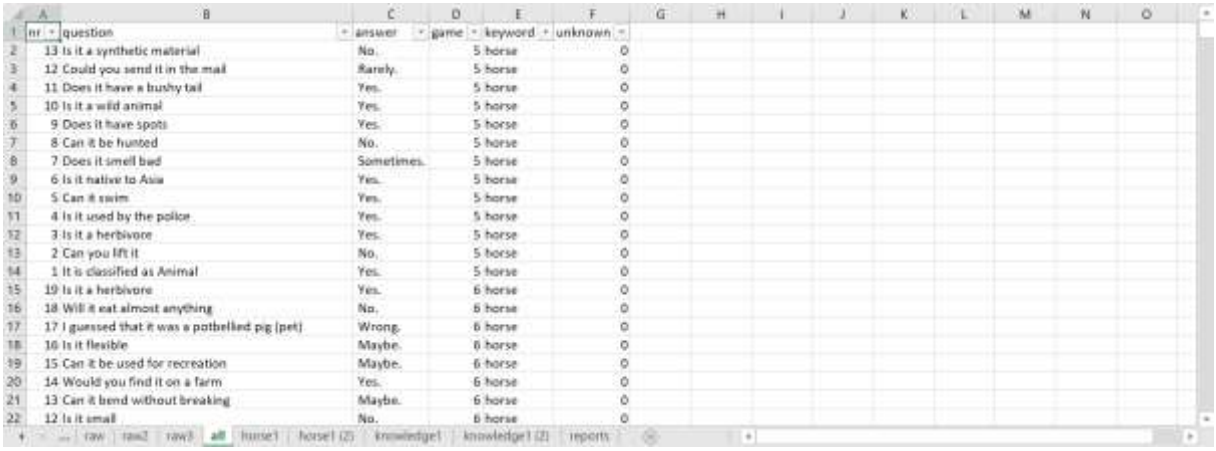
Figure Nr1 shows the steps of a game:

Figure Nr1: The process of an online 20Q-game (source: own presentation)

Data processing steps

If a lot of people play the game in the background with the same keyword (like horse or knowledge), then it is possible to see a lot of parallel guessing processes where the robot selects questions from the question pool in a randomized or even in a logical way. The same question can be answered from the players free – it means: quasi each human misunderstandings could be detected concerning the keyword and the particular question.

Based on a lot of games, a database can be defined (see Figure Nr2):



nr	question	answer	game	keyword	unknown
13	Is it a synthetic material	No.	5	horse	0
12	Could you send it in the mail	Rarely.	5	horse	0
11	Does it have a bushy tail	Yes.	5	horse	0
10	Is it a wild animal	Yes.	5	horse	0
9	Does it have spots	Yes.	5	horse	0
8	Can it be hunted	No.	5	horse	0
7	Does it smell bad	Sometimes.	5	horse	0
6	Is it native to Asia	Yes.	5	horse	0
5	Can it swim	Yes.	5	horse	0
4	Is it used by the police	Yes.	5	horse	0
3	Is it a herbivore	Yes.	5	horse	0
2	Can you lift it	No.	5	horse	0
1	It is classified as Animal	Yes.	5	horse	0
19	Is it a herbivore	Yes.	6	horse	0
18	Will it eat almost anything	No.	6	horse	0
17	I guessed that it was a potbellied pig (pet)	Wrong.	6	horse	0
16	Is it flexible	Maybe.	6	horse	0
15	Can it be used for recreation	Maybe.	6	horse	0
14	Would you find it on a farm	Yes.	6	horse	0
13	Can it bend without breaking	Maybe.	6	horse	0
12	Is it small	No.	6	horse	0

Figure Nr2: Structure and content of the database (source: own presentation)

The general database can contain information about

- the ranking number of a question within a game (nr)
- the question itself (question)
- the particular answers for the particular question in the particular game (answer)
- the id of the game (game)
- the keyword needing to guess (keyword)
- and arbitrary further columns like a binary status variable (answer is unknown or not)

Specialities making the integration of data (game by game) for a complex challenge:

- Each game can have different amount of questions.
- The questions having the answer of “unknown” do not have a ranking number within the particular game.
- The different basic parameters (like age, thinking/cultural basis, etc.) can lead to different output-formats: see
    - starting letter is a space or
    - not in case of a question or is a dot available as last letter in a string, etc.

Sooner or later, the consolidated data asset can be ensured. This data assets can be analysed based on the Excel-reporting wizard (pivot). Figure Nr3 shows the basic report about the keyword of “horse” where

- the rows are the question having more than one single mentioning
- the columns are the potential answers
- the cells contain the ratio of the chosen answer options compared to the sum of the row (=100%)

- the coloured background means: the more green is a cell the more simple could be answered the question

Figure Nr3: The basic pattern of the keyword of “horse” (source: own presentation)

The keyword of horse generates 171 questions. 68 questions were used at least two-times.

The same pattern for the keyword of knowledge can be seen in Figure Nr4:

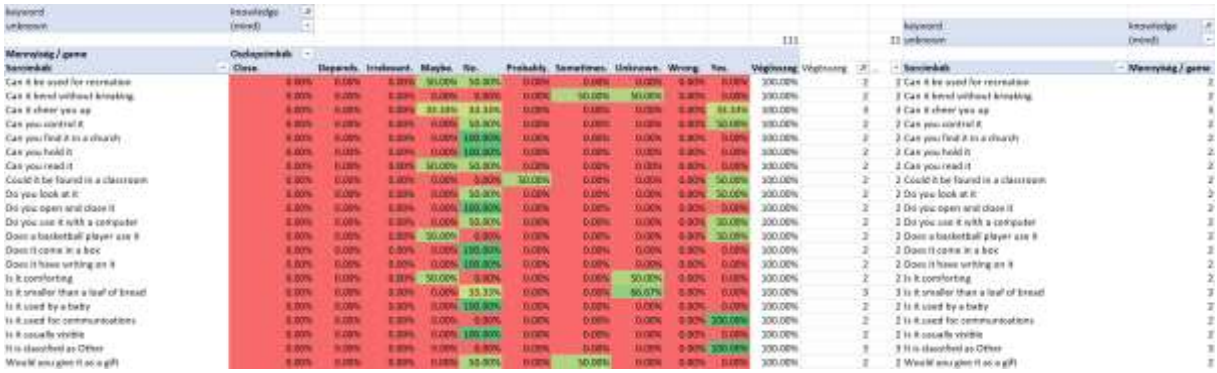


Figure Nr4: The basic pattern of the keyword of “knowledge” (source: own presentation)

The keyword of horse got selected in order to demonstrate a word with lower risk potential. The keyword of knowledge should present however a higher level of risks.

The basic patterns of the two words are to complex to derive/see relevant variables for risk-descriptions. Therefore, it is necessary to create new views (see Figure Nr5-6):

The relevant indicators can be seen on the bottoms of the figures:

- ratio1: 100% position for option “yes” + option “no” compared to the number of the visible questions
- ratio2: amount of more-times-used question compared to all question used in case of the keyword
- average: average using frequency of the visible questions

Further indicators could be:

- ratio of the affected answer-options: the average of the amounts of the used answer options in case of each visible questions compared to the permitted amount of answer-options (it is the same number for each questions)

- standard deviation of the using frequencies of the visible questions
- standard deviation of the ratios of the affected answer-options, …

The above listed potential indicators can be accepted as indicators for risk-estimations of the words in general because they have directions like:

- the higher is the value of ratio1 the lower is the interpretation-risk of a word
- the higher is the value of ratio2 the lower is the interpretation-risk of a word
- the higher is the value of the average frequency of the visible questions the lower is the interpretation-risk of a word
- the lower is the ratio of the affected answer-options the lower is the interpretation-risk of a word
- the lower is the standard deviation of the using frequencies of the visible questions the lower is the interpretation-risk of a word
- the lower is the standard deviation of the ratios of the affected answer-options the lower is the interpretation-risk of a word
- the lower is the ratio of unknown answer-option the lower is the interpretation-risk of a word (see Figure Nr7)
- the lower is average amount of the needed question for a correct guess the lower is the interpretation-risk of a word (see Figure Nr7)

keyword	horse					keyword	horse			
unknown	(mind)					unknown	(mind)			
			123		45					123
Mennyiség / game	Oszlopcímkék					Mennyiség / game	Oszlopcímkék			
Sorcímkék	No.	Yes.	Végösszeg			Sorcímkék	No.	Yes.	Végösszeg	
Can it be stolen	0.00%	100.00%	100.00%			Can it be stolen		2	2	
Can it climb	100.00%	0.00%	100.00%			Can it climb	3		3	
Can it help you find your way	0.00%	100.00%	100.00%			Can it help you find your way		2	2	
Can it jump	0.00%	100.00%	100.00%			Can it jump		2	2	
Can you find it in a house	100.00%	0.00%	100.00%			Can you find it in a house	2		2	
Can you hold it	50.00%	50.00%	100.00%			Can you hold it	1	1	2	
Can you lift it	100.00%	0.00%	100.00%			Can you lift it	2		2	
Can you see it in a zoo	0.00%	100.00%	100.00%			Can you see it in a zoo		2	2	
Do you hold it when you use it	50.00%	50.00%	100.00%			Do you hold it when you use it	1	1	2	
Does it dig holes	100.00%	0.00%	100.00%			Does it dig holes	2		2	
Does it have a bushy tail	25.00%	75.00%	100.00%			Does it have a bushy tail	1	3	4	
Does it have a horn	100.00%	0.00%	100.00%			Does it have a horn	3		3	
Does it have a long neck	66.67%	33.33%	100.00%			Does it have a long neck	2	1	3	
Does it have a long tail	0.00%	100.00%	100.00%			Does it have a long tail		4	4	
Does it have big ears	0.00%	100.00%	100.00%			Does it have big ears		2	2	
Does it have fangs	100.00%	0.00%	100.00%			Does it have fangs	3		3	
Does it have four legs	0.00%	100.00%	100.00%			Does it have four legs		3	3	
Does it have fur	33.33%	66.67%	100.00%			Does it have fur	1	2	3	
Does it have short fur	0.00%	100.00%	100.00%			Does it have short fur		2	2	
Does it have spots	50.00%	50.00%	100.00%			Does it have spots	1	1	2	
Does it help accomplish tasks	0.00%	100.00%	100.00%			Does it help accomplish tasks		2	2	
Does it hold water	100.00%	0.00%	100.00%			Does it hold water	3		3	
Does it jump	0.00%	100.00%	100.00%			Does it jump		2	2	
Does it live in the forest	100.00%	0.00%	100.00%			Does it live in the forest	2		2	
Does it provide protection	100.00%	0.00%	100.00%			Does it provide protection	2		2	
Does it require specific knowledge to use it	0.00%	100.00%	100.00%			Does it require specific knowledge to use it		2	2	
Is it a herbivore	0.00%	100.00%	100.00%			Is it a herbivore		4	4	
Is it a small mammal	50.00%	50.00%	100.00%			Is it a small mammal	1	1	2	
Is it a synthetic material	100.00%	0.00%	100.00%			Is it a synthetic material	2		2	
Is it a wild animal	33.33%	66.67%	100.00%			Is it a wild animal	1	2	3	
Is it awake at night	100.00%	0.00%	100.00%			Is it awake at night	2		2	
Is it considered intelligent	0.00%	100.00%	100.00%			Is it considered intelligent		2	2	
Is it dangerous	100.00%	0.00%	100.00%			Is it dangerous	3		3	
Is it furry	0.00%	100.00%	100.00%			Is it furry		4	4	
Is it heavy	0.00%	100.00%	100.00%			Is it heavy		4	4	
Is it larger than a microwave oven (or bread box)	0.00%	100.00%	100.00%			Is it larger than a microwave oven (or bread box)		2	2	
Is it small	100.00%	0.00%	100.00%			Is it small	3		3	
Is it smart	0.00%	100.00%	100.00%			Is it smart		2	2	
Is it spotted	100.00%	0.00%	100.00%			Is it spotted	4		4	
Is it striped	100.00%	0.00%	100.00%			Is it striped	3		3	
Is it used by the police	0.00%	100.00%	100.00%			Is it used by the police		2	2	
Is it worth a lot of money	0.00%	100.00%	100.00%			Is it worth a lot of money		2	2	
Will it eat almost anything	100.00%	0.00%	100.00%			Will it eat almost anything	2		2	
Would you find it on a farm	0.00%	100.00%	100.00%			Would you find it on a farm		4	4	
Would you use it daily	0.00%	100.00%	100.00%			Would you use it daily		2	2	
100%	16	21	45	82%	37%					2.56
				ratio1	ratio2					average

Figure Nr5: The characteristics of the keyword of “horse” (source: own presentation)

The above-listed indicators can be completed for each word in a common OA-matrix (see Figure Nr8) where objects are the words and the attributes are the indicators. Figure Nr6 demonstrates the

characteristics of the keyword of “knowledge”. It seems to be trivial, that the abstraction levels of words like “knowledge” should be higher than for words like “horse”. Therefore, it is to expect, that the indicators are able to reflect these relationships.

keyword	knowledge						keyword	knowledge		
unknown	(mind)						unknown	(mind)		
				91		13				91
Mennyiség / game	Oszlopcímkék						Mennyiség / game	Oszlopcímkék		
Sorcímkék	No.	Yes.	Végösszeg				Sorcímkék	No.	Yes.	Végösszeg
Can it cheer you up	50.00%	50.00%	100.00%				Can it cheer you up	1	1	2
Can you control it	50.00%	50.00%	100.00%				Can you control it	1	1	2
Can you find it in a church	100.00%	0.00%	100.00%				Can you find it in a church	2		2
Can you hold it	100.00%	0.00%	100.00%				Can you hold it	2		2
Do you look at it	50.00%	50.00%	100.00%				Do you look at it	1	1	2
Do you open and close it	100.00%	0.00%	100.00%				Do you open and close it	2		2
Do you use it with a computer	50.00%	50.00%	100.00%				Do you use it with a computer	1	1	2
Does it come in a box	100.00%	0.00%	100.00%				Does it come in a box	2		2
Does it have writing on it	100.00%	0.00%	100.00%				Does it have writing on it	2		2
Is it used by a baby	100.00%	0.00%	100.00%				Is it used by a baby	2		2
Is it used for communications	0.00%	100.00%	100.00%				Is it used for communications		2	2
Is it usually visible	100.00%	0.00%	100.00%				Is it usually visible	2		2
It is classified as Other	0.00%	100.00%	100.00%				It is classified as Other		3	3
100%	7	2	13	69%	14%					2.08
				ratio1	ratio2					average

Figure Nr6: The characteristics of the keyword of “knowledge” (source: own presentation)

keywords	average / nr of questions
horse	10
knowledge	13

keywords	Amount / unknown	Amount / question	ratio
horse	5	313	2%
knowledge	11	135	8%

Figure Nr7: Further characteristics of the words (source: own presentation)

direction	0	0	0				1	1	
OAM	ratio1	ratio2	average	...	...	...	amount of questions	ratio of unknown options	Y0
horse	82%	37%	2.56	...	...	...	10	2%	1000000
knowledge	69%	14%	2.08	...	...	...	13	8%	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000
further words...	...	...	...	...	...	...	...	...	1000000

Figure Nr8: Comparing words in frame of an antidiscriminative approach (source: own presentation)

The risk potential of the words should be estimated in frame of an antidiscriminative model (https://miau.my-x.hu/myx-free/coco/) where the objective is the prove whether each word can have the same evaluation value? If there are just two words where one of them has always a better position than the other one, then it is impossible to accept the same risk potential for each word – like here and now at once.

A “horse” has a lower abstraction potential than the keyword “knowledge”. This could also be demonstrated through the real guesses of the 20q.net-robot (see XLSX-object with detailed backgrounds):

- the “horse” could be approximated like
  - racehorse
  - pony
  - mustang,
  - Clydesdale (horse), etc.
- the “knowledge” could be approximated like
  - idea
  - logic
  - brain
  - concept, etc.

Unfortunately, the closeness to the basic keyword can not be measured in a direct way. But the ratio of the potential synonyms compared to the amount of the games (see: direction: the less - the lower risk) could also be used as a kind of synonyms.

Conclusions

As it can be seen, the 20Q-game is capable of exploring the potential of distortions of meanings. The frequently used question with more involved answer-options can be interpreted as a kind of misunderstanding potential where the answerers are not sure enough what should be answered. It could be focused on this kind of question in a more massive way (see Figure Nr3-4).

Parallel, it is possible to integrate status variables like age, sex, culture, country (region) into the general database. Therefore, it would also be possible to see regional, cultural, gender-oriented, gaining-oriented, etc. differences concerning the questions, answer-options, risk-indicators.

If the 20q.net-servers create and archive logs then the above-outlined steps could be processed at one!