

# Closing in on open-ended patient questionnaires with text mining

Prof. Irena Spasić  
i.spasic@cs.cardiff.ac.uk

Cardiff School of **Computer Science & Informatics**

<http://www.cs.cf.ac.uk>



1

wellcome trust

## Team

- Dr Kate BUTTON **Healthcare Sciences**
- Prof. Andy SMITH **Psychology**
- Mr David OWEN **Computer Science**



2

## Outline

- measurement of self-care
  - closed-ended vs. **open**-ended questionnaire
  - quantitative vs. **qualitative** analysis
- qualitative analysis based on **text mining**
  - **ordinal text classification** against a fixed, discrete rating scale, e.g. not at all < somewhat < a lot
  - **sentiment analysis**, e.g. negative < neutral < positive
  - **named entity recognition**
- results and next steps



3

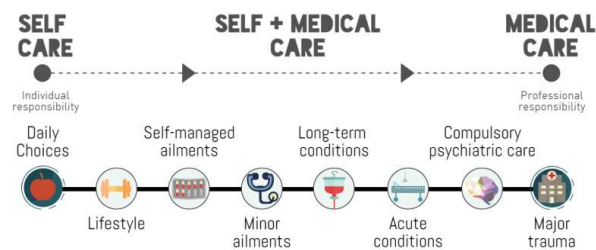
## Self-care



4

## Self-care

- all actions taken by people to recognise, treat and manage their own health
- independently or in partnership with the healthcare system



- "We need to fully understand ... what support people need for good self-care." NHS

5

## Knee injury & Osteoarthritis Outcome Score

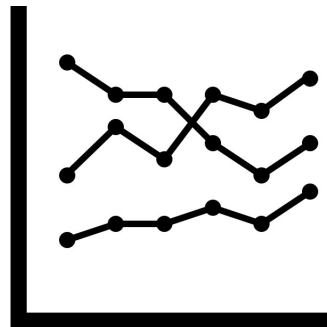
- an instrument for assessing patients' **opinion** about their knee and associated problems
- 42-item self-administered questionnaire  
<http://www.koos.nu/>
- ... to assess 5 outcomes: **(1)** pain, **(2)** other symptoms, **(3)** daily living, **(4)** sport & recreation, **(5)** quality of life
- Roos & Lohmander (2003) *Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis*. *Health and Quality of Life Outcomes*, 1:64.



6

## KOOS

- extensively used for clinical/research purposes
- used to monitor groups & individuals over time
- intended to be used **short-term**
  - to assess changes from week to week induced by treatment
- ... or **long-term**
  - over years following a primary injury or OA



## Pros & cons

- pros
  - user friendly
  - relatively quick to complete (approx. 10min)
- cons
  - closed-ended questions
  - answers limited to few options provided
  - 5-point Likert scale

Knee Injury and Osteoarthritis Outcome Score (KOOS), English version LK1.0

**KOOS KNEE SURVEY**

Today's date: \_\_\_/\_\_\_/\_\_\_ Date of birth: \_\_\_/\_\_\_/\_\_\_

Name: \_\_\_\_\_

**INSTRUCTIONS:** This survey asks for your view about your knee. This information will help us keep track of how you feel about your knee and how well you are able to perform your usual activities. Answer every question by ticking the appropriate box, only **one** box for each question. If you are unsure about how to answer a question, please give the best answer you can.

**Symptoms**  
These questions should be answered thinking of your knee symptoms during the last week.

S1. Do you have swelling in your knee?  
 Never  Rarely  Sometimes  Often  Always

S2. Do you feel grinding, hear clicking or any other type of noise when your knee moves?  
 Never  Rarely  Sometimes  Often  Always

S3. Does your knee catch or hang up when moving?  
 Never  Rarely  Sometimes  Often  Always

S4. Can you straighten your knee fully?  
 Always  Often  Sometimes  Rarely  Never

S5. Can you bend your knee fully?  
 Always  Often  Sometimes  Rarely  Never

**Stiffness**  
The following questions concern the amount of joint stiffness you have experienced during the last week in your knee. Stiffness is a sensation of restriction or slowness in the ease with which you move your knee joint.

S6. How severe is your knee joint stiffness after first waking in the morning?  
 None  Mild  Moderate  Severe  Extreme

S7. How severe is your knee stiffness after sitting, lying or resting later in the day?  
 None  Mild  Moderate  Severe  Extreme

## Limitations

- example

Q2. Have you modified your life style to avoid potentially damaging activities to your knee?				
Not at all	Mildly	Moderately	Severely	Totally
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q3. How much are you troubled with lack of confidence in your knee?				
Not at all	Mildly	Moderately	Severely	Extremely
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

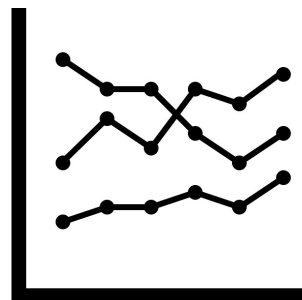
- we get answers to **what?**
- ... but not to **how?** and **why?**



9

## Limitations

- very limited insight into patient's **opinion**



- quantitative vs. qualitative research



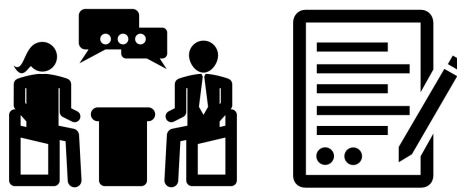
10

# Qualitative vs. quantitative analysis



## Qualitative research

- data collection
  - interviews
  - focus groups
  - questionnaires
- data analysis
  - content analysis
  - thematic analysis
- cons
  - sampling bias
  - low statistical power



Node Name	Count	Percentage	Occurrences	Occurrences	Occurrences	Occurrences
Free Nodes						
Nodes						
Interview	1	100%	1	1	1	1
Focus Group	1	100%	1	1	1	1
Questionnaire	1	100%	1	1	1	1
Other	1	100%	1	1	1	1



## Text analytics

- emerged as a potential solution to the problems of
  - information overload
  - information overlook
- uses techniques from
  - natural language processing
  - knowledge management
  - data mining



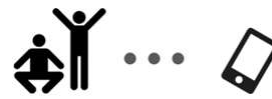
to efficiently process large text collections



13

## Our approach

- qualitative analysis... on steroids
- take advantage of modern technology to facilitate remote **interaction** with patients on a **regular basis**
- ... something traditional focus groups and interviews cannot provide easily
- patients report on their experience **online** with regular e-mail prompts
- ... then crunch the data using text analytics



14

## Data collection

- KLOG – a self-administered open-ended questionnaire
  - 10 questions
  - based on KOOS
- accessibility
  - secure online platform
  - Internet-enabled device
  - responsive mode



## Example: closed-ended question(s)

### Symptoms

These questions should be answered thinking of your knee symptoms during the **last week**.

S1. Do you have swelling in your knee?

Never  Rarely  Sometimes  Often  Always

S2. Do you feel grinding, hear clicking or any other type of noise when your knee moves?

Never  Rarely  Sometimes  Often  Always

S3. Does your knee catch or hang up when moving?

Never  Rarely  Sometimes  Often  Always

S4. Can you straighten your knee fully?

Always  Often  Sometimes  Rarely  Never

S5. Can you bend your knee fully?

Always  Often  Sometimes  Rarely  Never



## Example: open-ended question



[Home](#) [Log out](#)

Can you describe any knee **symptoms** you have experienced over the past week?

**Answer:**

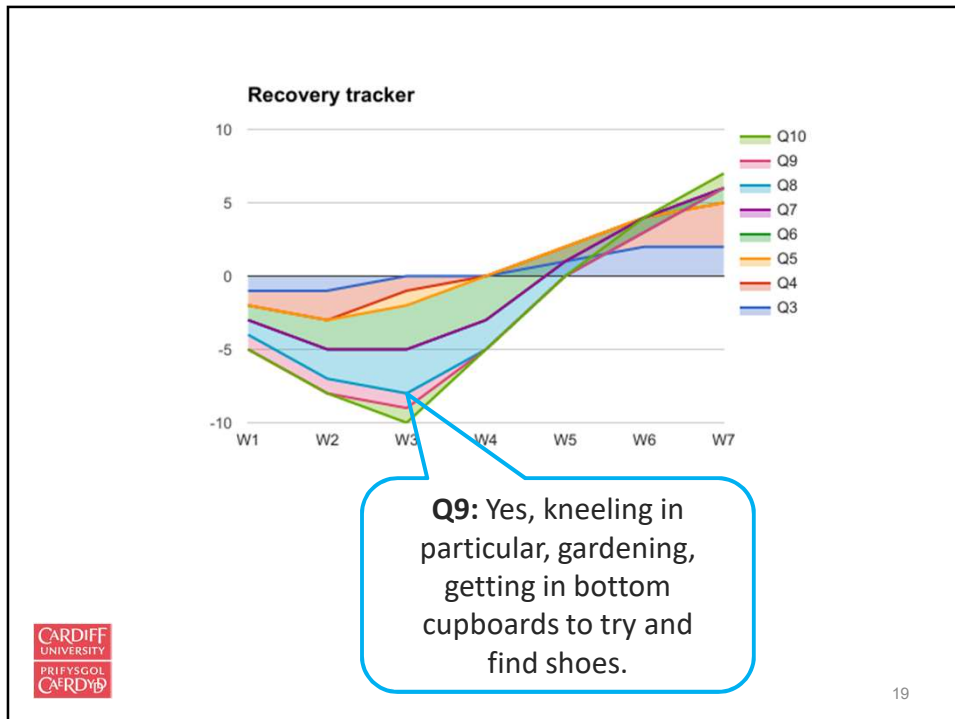
I am having trouble getting full extension as well. guess its more accurate to say, It takes a while. At first when I extend leg it hurts bad around the bottom of the knee cap after a few minutes the knee gets more extension and then after about 5 minutes it goes flat like the other one. This is new for me and appeared after I overdid walking and started doing some heel props on a chair at work to improve my extension. Surgeon says it was good but of a few degrees. This was all about 2 weeks ago, I have no swelling, just some bruising on the bottom of knee cap and a real sore spot to touch on the bottom inside of my knee just to the inside of the patella tendon. It was not like this when the surgeon examined me, now when physio asks me to straighten let it take a few minutes for it to stretch out and it hurts around the bottom of knee like I said.

Next



17

ID	Question	Coding
Q1	What knee <b>condition</b> are you currently receiving treatment for?	what? where?
Q2	What <b>treatment</b> are you currently using for your knee?	what?
Q3	Have there been any <b>changes</b> to your knee condition over the past week?	↓ worse < same < better ↑
Q4	How <b>confident</b> do you feel about looking after your knee?	↓ no < fairly < fully ↑
Q5	Can you describe any knee <b>stiffness</b> you have experienced over the past week?	↓ severe < some < none ↑
Q6	Can you describe any knee <b>pain</b> you have experienced over the past week?	↓ severe < some < none ↑
Q7	What knee <b>symptoms</b> other than stiffness and pain you have experienced over the past week?	↓ severe < some < none ↑
Q8	Has your knee condition limited your ability to carry out your <b>day to day tasks</b> over the past week?	↓ a lot < somewhat < no ↑
Q9	Has your knee condition limited your ability to carry out your <b>work, hobbies or exercise</b> over the past week?	↓ a lot < somewhat < no ↑
Q10	Do you have any other comments?	↓ neg < neutral < pos ↑ <sup>18</sup>



## Advantages

- therapeutic benefits (writing therapy)
- listening to patients
- clinical decision making

The diagram illustrates a workflow where a patient or user interacts with a digital tool (laptop) that communicates with a healthcare professional (doctor). The doctor's role involves listening to patients and making clinical decisions, represented by the checklist with a thumbs-up indicating a positive outcome.

CARDIFF UNIVERSITY  
PRIFYSGOL CARDIFF

20

## Closing in open-ended questions



21

ID	Question	Coding
Q1	What knee <b>condition</b> are you currently receiving treatment for?	what? where?
Q2	What <b>treatment</b> are you currently using for your knee?	what?
Q3	Have there been any <b>changes</b> to your knee condition over the past week?	↓ worse < same < better ↑
Q4	How <b>confident</b> do you feel about looking after your knee?	↓ no < fairly < fully ↑
Q5	Can you describe any knee <b>stiffness</b> you have experienced over the past week?	↓ severe < some < none ↑
Q6	Can you describe any knee <b>pain</b> you have experienced over the past week?	↓ severe < some < none ↑
Q7	What knee <b>symptoms</b> other than stiffness and pain you have experienced over the past week?	↓ severe < some < none ↑
Q8	Has your knee condition limited your ability to carry out your <b>day to day tasks</b> over the past week?	↓ a lot < somewhat < no ↑
Q9	Has your knee condition limited your ability to carry out your <b>work, hobbies or exercise</b> over the past week?	↓ a lot < somewhat < no ↑
Q10	Do you have any other comments?	↓ neg < neutral < pos ↑ <sup>22</sup>

## Unified Medical Language System (UMLS)

A service of the U.S. National Library of Medicine | National Institutes of Health

UMLS Terminology Services  
Metathesaurus Browser

Search: swelling  
Release: 2016AA  
Search Type: Word  
Source: All Sources

Search Results (3112)  
[ 1 - 25 ]

- C0013604 Edema
- C0038999 Swelling
- C0577599 Swelling absent
- C1411141 wandering swelling
- C0000731 Abdomen distended
- C0006114 Cerebral Edema
- C0006152 Swelling of breast
- C0026244 Mitochondrial Swelling
- C0035333 Retinitis
- C0151602 Facial swelling

Concept [C0013604] Edema

Semantic Types  
Finding [T033]

Definitions  
CSP | swelling from excessive accumulation of serous fluid in tissue.  
HPO | An abnormal accumulation of fluid beneath the skin, or in one or more cavities of the body. [HPO:probinson]  
MEDLINEPLUS |  
Edema means swelling caused by fluid in your body's tissues. It usually occurs in the feet, ankles and legs, but it can involve your entire body.  
Causes of edema include

- Eating too much salt
- Sunburn
- Heart failure
- Kidney disease
- Liver problems from cirrhosis
- Pregnancy
- Problems with lymph nodes, especially after mastectomy
- Some medicines

23

## Semantic types

A service of the U.S. National Library of Medicine | National Institutes of Health

UMLS Terminology Services  
Semantic Network Browser

2016AB

(A) Entity

- (A1) Physical Object
  - (A1.1) Organism
  - (A1.2) Anatomical Structure
  - (A1.3) Manufactured Object
  - (A1.4) Substance
- (A2) Conceptual Entity
  - (A2.1) Idea or Concept
  - (A2.2) Finding
    - (A2.2.1) Laboratory or Test Result
    - (A2.2.2) Sign or Symptom
  - (A2.3) Organism Attribute
  - (A2.4) Intellectual Product
  - (A2.5) Language
  - (A2.6) Occupation or Discipline
  - (A2.7) Organization
    - (A2.8) Group Attribute
  - (A2.9) Group
- (B) Event
- (H) isa
- (R) associated\_with

Sign or Symptom

Definition  
An observable manifestation of a disease or condition based on clinical judgment, or a manifestation of a disease or condition which is experienced by the patient and reported as a subjective observation.

Properties  
Unique Identifier: T184  
Tree Number: A2.2.2

Parents  
Finding

Relations  
Inverse Relations  
Inherited Relations  
Inverse Inherited Relations

24

ID	Question	Coding
Q1	What knee <b>condition</b> are you currently receiving treatment for?	what? where?
Q2	What <b>treatment</b> are you currently using for your knee?	what?
Q3	Have there been any <b>changes</b> to your knee condition over the past week?	↓ worse < same < better ↑
Q4	How <b>confident</b> do you feel about looking after your knee?	↓ no < fairly < fully ↑
Q5	Can you describe any knee <b>stiffness</b> you have experienced over the past week?	↓ severe < some < none ↑
Q6	Can you describe any knee <b>pain</b> you have experienced over the past week?	↓ severe < some < none ↑
Q7	What knee <b>symptoms</b> other than stiffness and pain you have experienced over the past week?	↓ severe < some < none ↑
Q8	Has your knee condition limited your ability to carry out your <b>day to day tasks</b> over the past week?	↓ a lot < somewhat < no ↑
Q9	Has your knee condition limited your ability to carry out your <b>work, hobbies or exercise</b> over the past week?	↓ a lot < somewhat < no ↑
Q10	Do you have any other comments?	↓ neg < neutral < pos ↑

## Features

**Q9:** Has your knee condition limited your ability to carry out your **work, hobbies or exercise** over the past week?

**A:** Not been able to carry out my normal job at work due to my injury as I am a commercial vehicle mechanic. Limited exercise and hobbies. I am very frustrated that I haven't been able to play squash due to the injury and golf as I can't twist my knee fully yet.

**C:** a lot < somewhat < no ?



## Features

**Q9:** Has your knee condition limited your ability to carry out your **work, hobbies or exercise** over the past week?

**A:** Not been able to carry out my normal **job** at **work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. I am very frustrated that I haven't been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.



**C:** a lot < somewhat < no ?

27

## Features

**Q9:** Has your knee condition limited your **ability** to carry out your **work, hobbies or exercise** over the past week?

**A:** Not been able to carry out my normal **job** at **work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. I am very frustrated that I haven't been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.



**C:** a lot < somewhat < no ?

28

## Features

**Q9:** Has your knee condition limited your **ability** to carry out your **work, hobbies or exercise** over the past week?

negation

**A:** **Not** been able to carry out my normal **job** at **work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. I am very frustrated that I **haven't** been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.



**C:** a lot < somewhat < no ?

29

## Features

**Q9:** Has your knee condition limited your **ability** to carry out your **work, hobbies or exercise** over the past week?

sentiment

**A:** Not been able to carry out my normal **job** at **work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. **I am very frustrated** that I haven't been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.



**C:** a lot < somewhat < no ?

30

<p><b>Features</b></p> <ul style="list-style-type: none"> <li>▪ concepts             <ul style="list-style-type: none"> <li>▪ frequency</li> <li>▪ context                 <ul style="list-style-type: none"> <li>▪ negation</li> <li>▪ sentiment</li> </ul> </li> </ul> </li> </ul>	<p><b>... methods</b></p> <ul style="list-style-type: none"> <li>... ontology mapping</li> <li>... dependency parsing</li> <li>... sentiment analysis</li> </ul>	<p><b>... tools</b></p> <ul style="list-style-type: none"> <li>... MetaMap</li> <li>... Core NLP</li> <li>... Core NLP</li> </ul>
--	--	---

free-text answers
feature vectors
predicted classes

feature extraction
classification model

31

### Supervised machine learning

**training**

feature vectors	class labels

→ machine learning algorithm

**classification**

feature vector	unknown class
	N/A

→ classification model
← predicted class

32



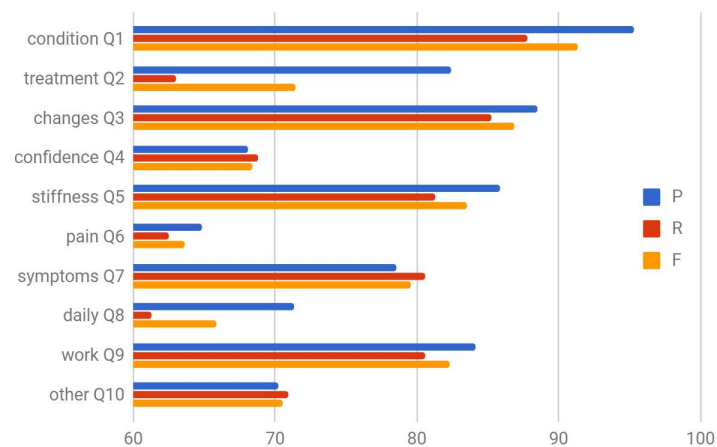
## Cross-validation results

Question	Topic	Method	Features	P (%)	R (%)	F (%)
Q1	condition	MetaMap	N/A	95.3	87.8	91.4
Q2	treatment	MetaMap	N/A	82.4	63.0	71.4
Q3	changes	complement naive Bayes	8	88.5	85.3	86.9
Q4	confidence	random tree	8	68.1	68.8	68.4
Q5	stiffness	logistic model tree	8	85.9	81.3	83.5
Q6	pain	complement naive Bayes	10	64.8	62.5	63.6
Q7	symptoms	naive Bayes	5	78.6	80.6	79.6
Q8	daily	ordinal class classifier	14	71.3	61.3	65.9
Q9	work	J48 pruned tree	14	84.1	80.6	82.3
Q10	other	Stanford Core NLP	N/A	70.2	70.9	70.5



33

## Cross-validation results



34

## Conclusion



35

## Conclusion

- Can we measure self-care with text analytics?
  - Absolutely!
- next steps
  1. get more training/test data
  2. re-train and properly evaluate on unseen data
  3. embed automated measurement into patient dashboard
- future work
  1. validate as an instrument
  2. explore speech-to-text for data collection



36