5️⃣

Interfaces

Course	RecSys
Files	Lecture3.5 - Interfaces.pdf
Link	
Status	Done
Type	Lecture

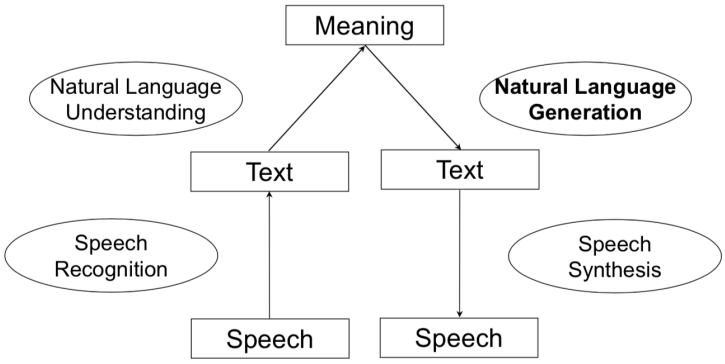
Explanation Metrics

- Transparency - how the system works
- Effectiveness - better decisions
- Trust - more confidence
- Persuasiveness - convincing
- Satisfaction - ease of use/enjoyment
- Scrutability - adjustable/modifiable
- Efficiency - faster decisions

- Justification: why a decision is good, without how it was made
- Interpretability is the degree to which a human can
  - understand the result

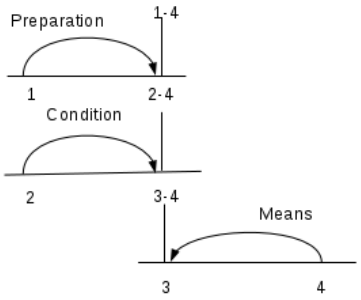
- consistently predict the models results

Natural Language Generation



- Gricean Maxims
  - Quantity - required info
  - Quality - true
  - Relevance - be relevant
  - Manner - be clear
- How?
  1. **Document planning**
     Content and structure
  2. **Microplanning**
     Words and sentences to use
  3. **Realisation**

     Put it together taking into account grammar
- Text Structuring
  - RST (Rhetorical Structure Theory) - kind of a way to represent grammar for a computer



  - Schemata: hand-crafted structuring rules
- Sentence aggregation
  - If a sentence keeps getting repeated, group it somehow instead of repeating
- Realiser systems - take care of a bunch of stuff like adjective ordering, spacing, punctuation, tenses, negation
  - simplenlg – relatively limited functionality, but well documented, fast, easy to use, tested
  - KPML – lots of functionality but poorly documented, buggy, slow
  - openccg – somewhere in between

Explanations for group recommendations

	Strategy	Explanation template
Consensus-based	*ADD*	*"Item X has been recommended to the group since it achieves the highest total rating.* ***This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."***
	FAI	***"The preference of user u_a was not considered in the last n decisions. Therefore, in this decision, item X has been recommended to the group since he/she likes it the most."***
Majority-based	*APP*	*"Item X has been recommended to the group since it achieves the highest number of ratings which are above a threshold θ.* ***This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."***
Borderline	*LMS*	*"Item X has been recommended to the group since no group member has a real problem with it.* ***This decision supports the preferences of users u_a and u_b who were treated less favorably in the last n decisions."***
	MAJ	*"Item X has been recommended to the group since most group members like it.* ***This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."***
	MPL	*"Item X has been recommended to the group since it achieves the highest of all individual group members' ratings.* ***This decision supports the preference of user u_a who was treated less favorably in the last n decisions."***

NLP Evaluation

- As a whole - does it work? (criteria could be those described in explanation metrics)

- Layered evaluation - break the system into components and evaluate each component
- Task Performance - does the system help achieve the communication goal?
- Task-based - "most respected", very expensive and time consuming, evaluation of a specific system, not an algorithm or idea
- Human ratings - most common, usually Likert scale, usually using crowd sourcing (Mechanical Turk)

Which is most meaningful?

	Task (extrinsic)	Ratings (intrinsic)
Real-world	**Most meaningful**	intermediate
Laboratory	intermediate	*Least meaningful*

Which is most costly?

	Task (extrinsic)	Ratings (intrinsic)
Real-world	**Most expensive**	intermediate
Laboratory	intermediate	*Cheapest*

- Metric - compare the output to "gold standard" (human written), limited confidence