

# NTCIR-11 Core Task: Spoken Query and Spoken Document Retrieval (SpokenQuery&Doc)

Tomoyosi Akiba (Toyohashi University of Technology)

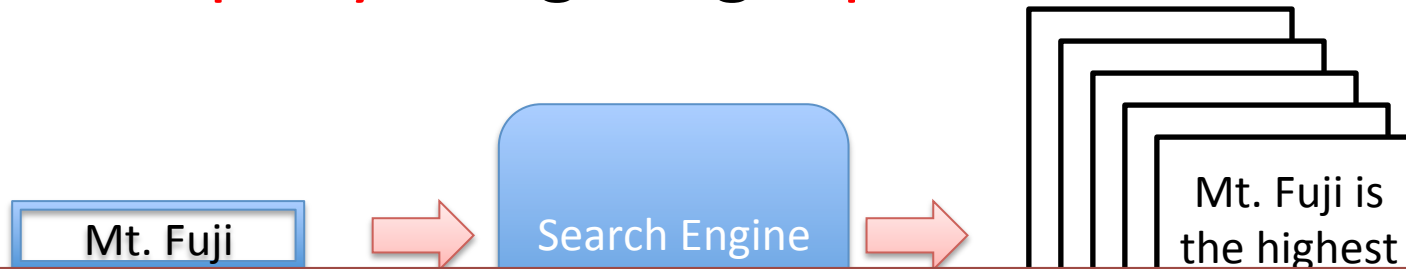
Hiromitsu Nishizaki (Yamanashi University)

Hiroaki Nanjo (Ryukoku University)

Gareth Jones (Dublin City University)

# Focus of SpokenQuery&Doc

- Information Retrieval by using “spontaneously spoken query” targeting “spoken documents”.



Overcome it by using Speech!



I'm going to go climbing **Mt. Fuji** in next summer vacation. Well, I hear the accident um... I hear the news about the accident in the climbing. Well, someone has missed or someone lost one's way in the mountain. So I would like to know what should I prepare for climbing, what equipment, equipment is necessary. I would like to know what should I prepare when I go climbing.

# Example of Spontaneously Spoken Query

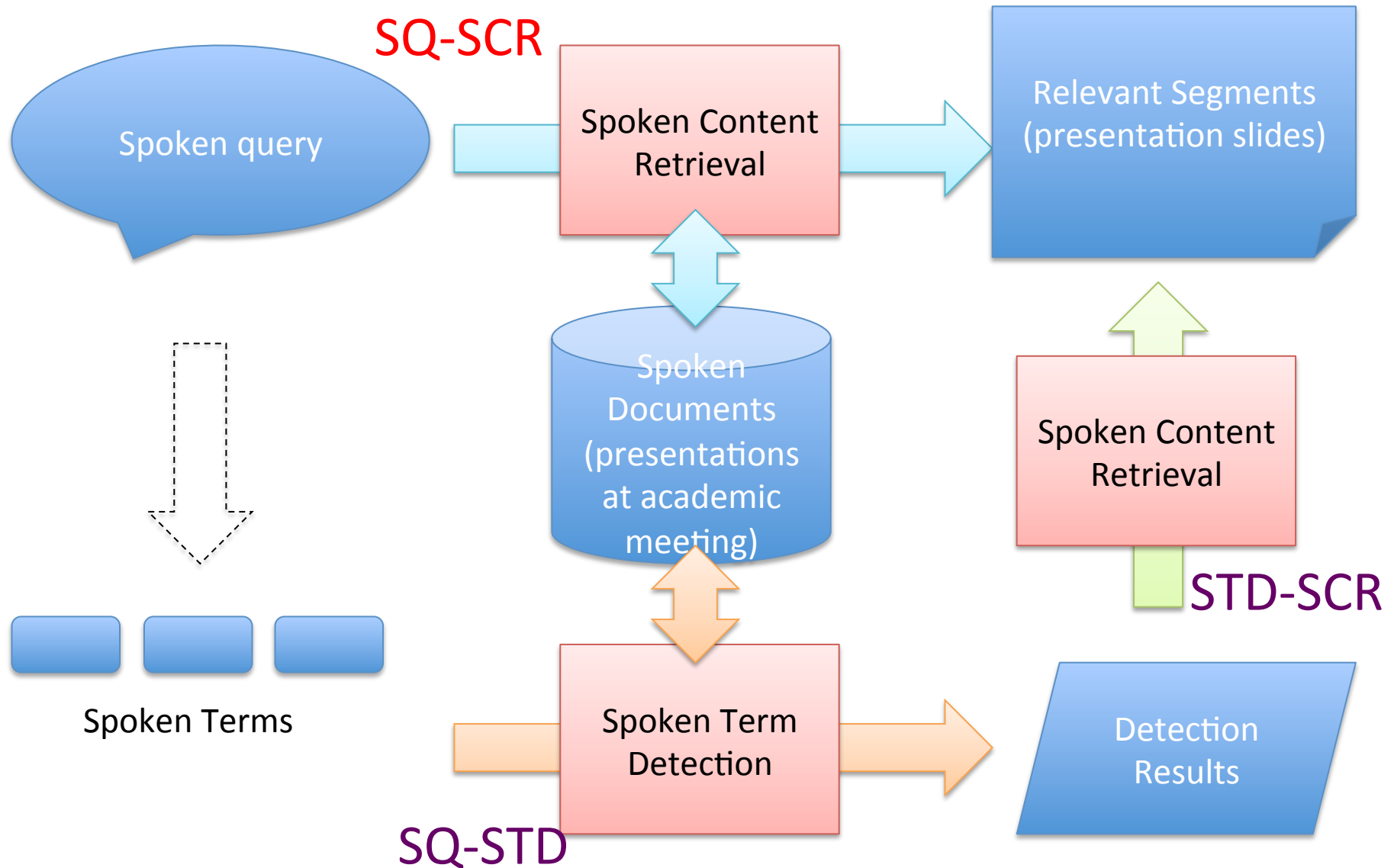
今年の夏休み(summer vacation)に始めて山(mountain)に登山に行くことになったんですけども。あの一登山は結構事故(accident-1)があの一やはり夏(summer)になるとよくニュース(new)で聞きますし、ま誰々が行方不明(go missing)になったとか遭難(accident-2)したとかそういう話が色々あると思うんですけども。あの一山(mountain)に登る(climb)時にはどういった心構え(keep in mind)と言うか、あの一装備(equipment)こういう装備が必要だとかこういうものがあるといいよ、とかそういったあの一山登り(mountain climbing)に関しての、山(mountain)に登る(climb)時についての心構え(keep in mind)について知りたいです

- PROS: provides rich clues for retrieval.
- CONS: unclear, ambiguous, redundant, error-prone

# Challenges in SpokenQuery&Doc

- How to get benefit from spontaneously spoken queries?
  - Make use of long queries.
  - Disambiguation.
  - Improve ASR for spontaneous speech.
  - Development of specific techniques for such IR.
- Spoken document retrieval (continued)
  - Evaluate spoken term detection (STD) and spoken content retrieval (SCR) at the same time by using common (spoken) query topics.
  - Text queries are also provided (as in SpokenDoc-1 and 2.)

# Task Design



# Why participate in SpokenQuery&Doc

- Participants will be provided many resources for IR and spoken language processing research.
  - Real recordings of spontaneously spoken queries.
  - Real recordings of academic meeting lectures.
  - Their rich transcriptions.
    - manual.
    - automatic (with various conditions.)
    - alignment with presentation slides.
- IR and NLP researchers are also welcomed.
  - Easy to apply your existing techniques to provided text transcriptions of speech.
- Web site will be open soon.
  - <http://www.nlp.cs.tut.ac.jp/ntcir11/>