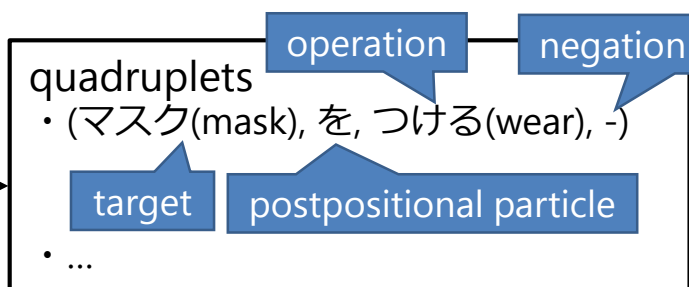# University of Hyogo at NTCIR-11 TaskMine by Dependency Parsing

Takayuki Yumoto (Graduate School of Engineering, University of Hyogo, Japan)

Oveview
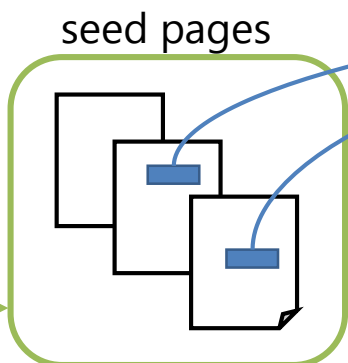
input : query
・花粉症対策をする (ease hay fever)

→ Web →

quadruplets
operation   negation
・(マスク(mask), を, つける(wear), -)
target   postpositional particle
・...

→

output : subtasks
・マスクをつける (wear a mask)
・目薬を使う(use a eyewash)
・...

1. Collection

sources = seed pages + detailed pages

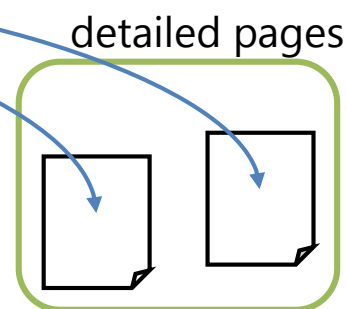(1) Collecting seed pages by query expansion

query string + "方法(method)"

↓

Search Engine → seed pages

(2) Collecting detailing pages

conditions:
・same domain
・all query keywords in anchor text (and a title of the seed page)
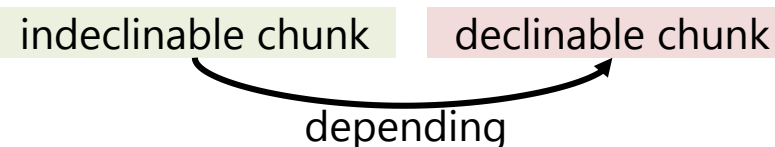
→ detailed pages

2. Extraction

We use a dependency parsing (Cabocha)

(1) Finding operation (and negation)

1. the end of sentence is the first candidate

マスク(mask)を / つけて(wear) / 出かける(go out)

2. the declinable chunk depending to the other candidates is also a candidate

indeclinable chunk   declinable chunk
depending

外に(outside) / 干さない(don't hang)

negation is extracted from a candidate chunk if it exists

(2) Finding target and postpositional particle

the indeclinable chunk depending to operation

マスク(mask)を / つけて(wear) / 出かける(go out)

3. Ranking

We compared two ranking methods
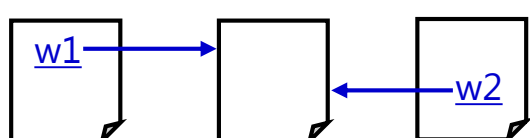
(A) Site frequency of quadruplets : Order by

(マスク(mask),を,つける(wear), -)

(B) Site frequency of target and operation

Order by Min(,), Max (,)

(マスク(mask),を, つける(wear), -)

Synonyms by Wikipedia corpus

w1 →   ← w2

w1 and w2 are regarded as synonyms

Result

given : total 50 queries of 4 categories

detail

	nDCG@1	nDCG@5	nDCG@10	nDCG@50
(A)	0.109	0.150	0.171	0.191
(B)	0.098	0.119	0.132	0.166
baseline	0.013	0.040	0.053	0.096

Category	nDCG@1	nDCG@5	nDCG@10	nDCG@50
Health	0.140	0.141	0.173	0.191
Education	0.000	0.075	0.107	0.161
Daily life	0.167	0.153	0.150	0.154
Sequential	0.100	0.237	0.253	0.259

Problems

- the part where subtasks are extracted
  e.g. the page describing methods to prevent a cold and mechanism to catch a cold

- same subtask in different expressions
  e.g. "wear a mask" = "use a mask"

- limitation of model : multiple targets are sometimes needed in a single subtask