OMP 지식구축 Stand Alone 매뉴얼

1. 환경설정

지식구축 모듈은 NLP모듈을 사용하기 때문에 knowledgeBuilding.jar는 고정된 위치에 존재 해야 한다. 만일 이것을 이것의 위치를 수정하고자 할 때 ./resource 와 ./config를 knowledgeBuilding.jar 가 존재하는 곳에 복사를 해야 한다.

지식구축기 (Knowledge Building)를 실행하기 위해서는 [예1]과 같이 [baseDirectory], [Topic Name], [encoding : UTF-8, MS949] 의 3가시 argument를 반드시 입력해야만 한다.

1. java –cp ./knowledgeBuilding.jar

com.skplanet.omp.knowledgeBuilding.KnowledgeBuildingTransaction\_StandAlone [baseDirectory] [Topic Name] [encoding : UTF-8, MS949]

[baseDirectory]는 지식구축 시 필요한 input(분석 대상 원문), output 및 분석을 위해 필요한 사전 정보를 가지는 최상위 디렉터리를 의미한다. 분석을 수행하기 위해서는 [예2]와 같이 전 [baseDirectory]와 4개의 서브 디렉터리([baseDirectory]/DATA, [baseDirectory]/ATTR, [baseDirectory]

/DICT, [baseDirectory]/EXPR)와 /EXPR의 서브 디렉토리 2개([baseDirectory]/EXPR/NEW, [base

Directory]/EXPR/OLD)가 반드시 존재해야 한다. 또한 [baseDirectory]/DICT 에는 불용어사전(stopword.txt), 일반표현사전(general.txt), 동의어사전(synonym.txt), 속성을 추출하기 위해 사용되는 Clue사전(clue.txt)가 반드시 존재해야 한다.

1. [baseDirectory]

[baseDirectory/DATA]  
[baseDirectory/ATTR]

[baseDirectory/DICT]

[baseDirectory/EXPR]

[baseDirectory/EXRP/NEW]

[baseDirectory/EXRP/OLD]

[Topic Name]은 분석결과가 저장될 때 [Topic Name].txt 와 같이 Naming 되어 저장된다.

[encoding : UTF-8, MS949]은 입력되는 분석대상 원문의 인코딩 형태를 의미 한다.

1. 실행

지식구축기가 존재하는 디렉터리에서 다음과 같이 지식구축기(knowledgeBuilding.jar)를 실행한다.

> java -cp ./knowledgeBuilding.jar com.skplanet.omp.knowledgeBuilding.KnowledgeBuildingTransaction\_StandAlone [baseDirectory] [Topic Name] [encoding : UTF-8, MS949]

1. 분석결과
   1. 속성 후보 추출 결과

속성 후보 추출 결과는 clue기반, 표현기반으로 분리되어 저장되며 각각[baseDirectory]/ATTR/[Topic Name]\_clue.txt과 [baseDirectory]/ATTR/[Topic Name]\_clue.txt 의 형태로 저장된다.

* 1. 속성-표현 후보 추출 결과

속성-표현 후보 추출 결과는 일반표현을 기반으로 긍부정도가 마킹된 것과 그렇지 않은clue기반, 것으로 분리되어 저장되며 각각 [baseDirectory]/EXPR/NEW/[Topic Name].txt과 [baseDirectory]/EXPR/NEW/[Topic Name]\_missed.txt 의 형태로 저장된다.

또한, 재분석을 고려하여 [baseDirectory]/EXPR/OLD/ 아래 동일한 형태로 속성-표현 후보가 저장되며 다음 분석 시 [baseDirectory]/EXPR/OLD/ 에 존재하는 표현의 경우 [baseDirectory]/EXPR/NEW/[Topic Name]\_conflict.txt 와 [baseDirectory]/EXPR/NEW/[Topic Name]\_missed\_conflict.txt의 형태로 저장된다.