

Lab Works

A design of web based Lab note system
using Information gain and Ontology

Sohn Jong-Soo
mis026@korea.ac.kr

2007.12.04



IIS Lab. Dept. of Computer Information Science Korea Univ.

0. 요약

■ Lab note

- 연구계획, 과정, 성과의 저장소
- 연구실내에서 지식의 전수

■ Lab Works

- 온톨로지 기술을 이용
 - 연구실 업무 및 학습활동을 분류
 - 학술용어 및 연구도메인을 체계화
- Information Theory를 이용
 - 가장 사용자에게 유용한 순서대로 글을 선택하고 정렬

1. Lab note

■ Lab note의 필요성

- Lab note를 사용하지 않는 경우
 - 연구지식이 전수되지 않음
 - 특정 문제를 해결함에 있어서 노력의 반복
 - 이전 연구의 진행과정 및 결론 도출 과정의 파악이 쉽지 않음
 - 새로운 연구원이 들어오는 경우
 - 진행되어온 연구에 관한 교육의 문제 발생
- 기존 해결방법
 - 일일 연구일지 작성
 - 연구실에서 발간된 논문 및 연구보고서 참고
 - 중간 결과물의 백업본 참고
 - 전화나 이메일을 통한 선배들의 조언

1. Lab note

■ Lab note의 개념

- 연구 계획, 과정, 성과의 저장소
 - Laboratory note = experiment data + idea + knowhow
 - 특정 프로젝트의 이력(history) 추적
 - 연구자간 연구결과 공유 및 토론의 수단
- 연구실내에서의 지식의 전수
 - 연구실의 모든 정보의 집합
 - 후발연구자가 빠른 시간 안에 연구실에 축적된 지식, 노하우를 익혀 후속연구를 진행할 수 있게 하는 도구
 - 연구실만의 노하우가 전수
 - 연구실 연구방법의 독자성 유지

1. Lab note

■ Lab note의 활용

- 협업 프로젝트의 수행 효율성
 - 다수의 연구자가 동일 프로젝트를 수행
 - 다른 연구자의 결과물에 대한 이해가 수월
 - 문제접근방법에 대한 토의가 쉽고 빠름
- 연구 독자성 증명 수단
 - 유사연구에 대해 독자성을 증명할 수 있는 근거
 - 공동연구의 경우 결과물에 대한 독자개발의 증명 시
 - 선 발명의 판단 근거
 - Conception 및 reduction to practice의 시점으로 발명자를 가리는 원칙 - 미국 특허법 : 선발명 주의 中

1. Lab note

■ Lab note의 활용 (계속)

- 지식 습득의 효율성 보장
 - 흡사한 분야의 선행연구를 거꾸로 추적하여 직면한 문제를 해결하기 위한 방법 고안이 쉬움
- 연구 지식의 축적
 - 프로젝트나 논문에 관한 연구 지식을 축적
 - 시간이 지날수록 더 많은 지식을 축적가능
 - 연구단절의 가능성을 낮춤

※ See also

[Advice on keeping a laboratory notebook](#)

[keeping a laboratory notebook](#)

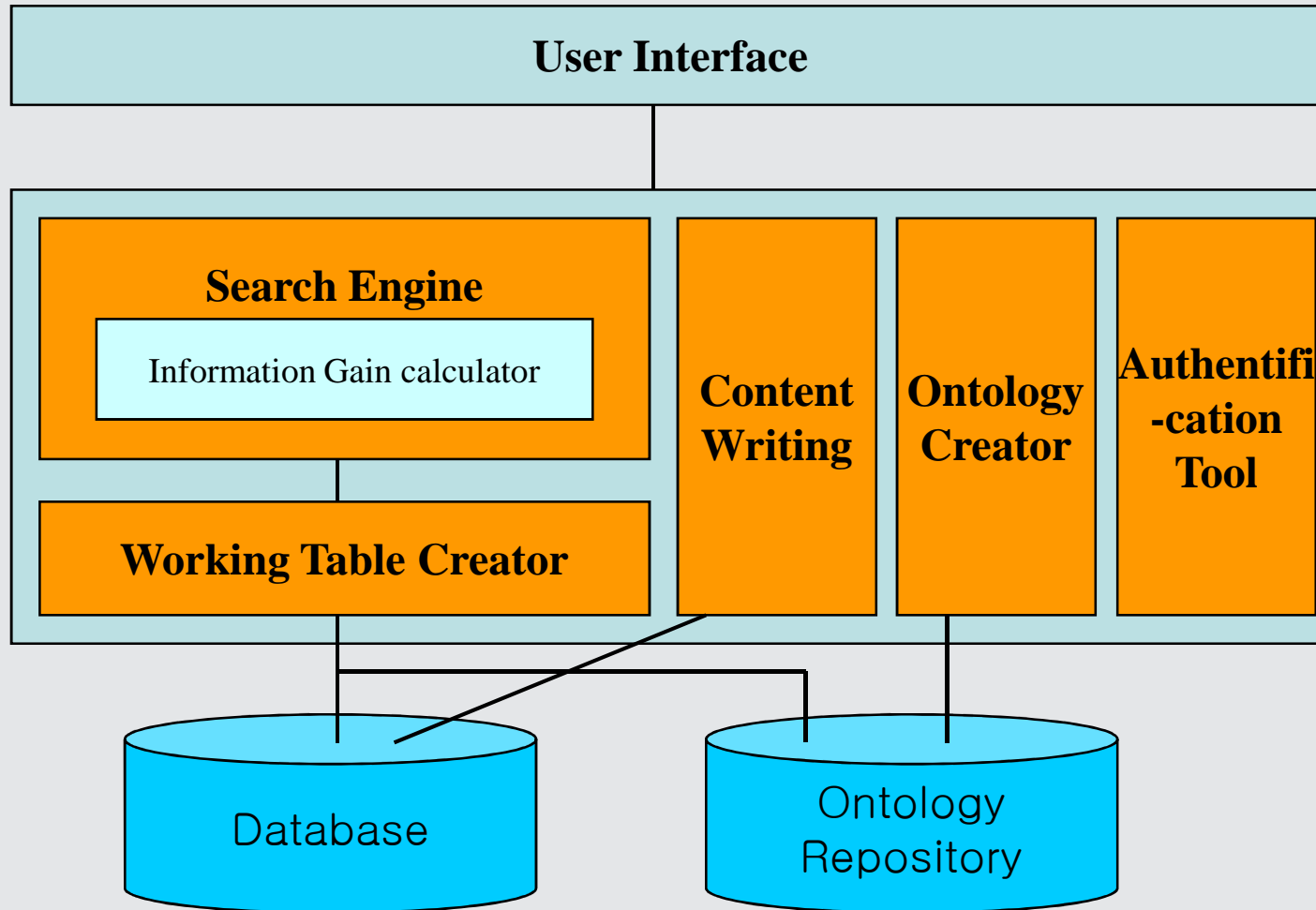
[Laboratory Notebooks](#)

1. Lab note

■ Lab note 작성 내용

- 1) Title and date
- 2) Members who assisted you, if any
- 3) Objectives
- 4) Procedure
- 5) References (Line, page, why .. Etc)
- 6) Date/results
- 7) Conclusion/discussion
- 8) Signatures/date

2. System architecture



2. System architecture

No.	Date	Name	Category	Subcategory	Keyword	Subject	content
1	07.11.30	손종수	Course Work	07Datamining	Information Theory, ID3	Gain(A) 구하기	...
2	04.10.02	이지윤	Paperwork	Background Study	Information Theory	이론학습..	...
3	05.05.11	이윤수	Course Work	05기계 학습	PCA, ANN	차원압축	...
4	04.04.27	이윤수	Course Work	04인공신경망	MLPBP, ANN	패턴인식	...
5	07.12.03	공신조	TA	인공지능 (Under)	ID3	조교수업준비	...
6	07.12.06	공신조	TA	자료구조	Graph, Search	조교수업준비	...

3. Category

CourseWork	학위 과정
Project	프로젝트
TA	조교
HumanRelationship	인간관계
ProfessorsWork	교수님 업무
PaperWork	논문관련연구
Lecture	강의
PaperReading	읽은 논문 정리
Terms	각종 용어 및 개념 정리
etc	기타

3. Category

CourseWork				
	개설년도			
		교과목명		
		교과과정		
		숙제 및 과제		
		시험 및 기말과제		
			중간시험	
			기말시험	
			기말과제	
				1차발표
				2차발표
		교과서		

3. Category

Terms			
	algorithm		
		Search	
		Clustering	
		Reasoning	
			ID3
	Theory		
		Information Theory	
		Graph	
		Tree	
	Language		
		C	
			IF
	Application		
	Domain		
		Datamining	
		Artificial Intelligence	
		Network	
		WebSystem	
		Security	

3. Category

■ 온톨로지 기술의 예

```
<owl:Class rdf:ID="GraphTheory">  
  <rdfs:subClassOf rdf:resource="Theory"/>  
  <owl:unionOf>  
    <owl:Class rdf:resource="#directedGraph"/>  
    <owl:Class rdf:resource="#undirectedGraph"/>  
  </owl:unionOf>  
</owl:Class>
```

- 온톨로지는 기본적으로 Tree구조를 갖지만 클래스와 클래스간의 관계와 속성을 기술하기 때문에
 - 온톨로지 병합 및 참조를 하다보면 Graph의 형태가 됨
- 따라서 카테고리 작성시엔 상 하위 클래스 관계만을 보며 검색할 땐 관계와 속성을 확인하여 정확한 검색을 유도

4. Content

No.	Date	Name	Category	Subcategory	Keyword	Subject	content
1	07.11.30	손종수	Course Work	07Datamining	Information Theory, ID3	Gain(A) 구하기	...
2	04.10.02	이지윤	Paperwork	Background Study	Information Theory	이론학습..	...
3	05.05.11	이윤수	Course Work	05기계 학습	PCA, ANN	차원압축	...
4	04.04.27	이윤수	Course Work	04인공신경망	MLPBP, ANN	패턴인식	...
5	07.12.03	공신조	TA	인공지능 (Under)	ID3	조교수업준비	...
6	07.12.06	공신조	TA	자료구조	Graph, Search	조교수업준비	...

5. Working table

No.	Cat	subCat	Upnode	Node	Equiv	Disjoint	One of	Intersec
1-1	Course	DM	Algo.	ID3	Null	Null	Null	InfTh.
1-2	Course	DM	Theory	InfTh.	Null	Null	Null	ID3
2-1	PaperW	BackStd	Theory	InfTh.	Null	Null	Null	ID3
3-1	Course	ML	Algo.	PCA	Null	Null	ANN	Null
3-2	Course	ML	Domain	ANN	Artificia..	Null	Null	Null
4-1	Course	ANN	Algo.	MLPBP	EBPA	Null	ANN	Null
4-2	Course	ANN	Domain	ANN	Artificia..	Null	Null	Null
5-1	TA	AI(U)	Algo.	ID3	Null	Null	Null	InfTh.
6-1	TA	DS(U)	Theory	Graph	Null	Null	Null	Search
6-2	TA	DS(U)	Algo.	Search	Null	Null	Null	Graph

6. 검색

■ Information Theory로 검색한 경우

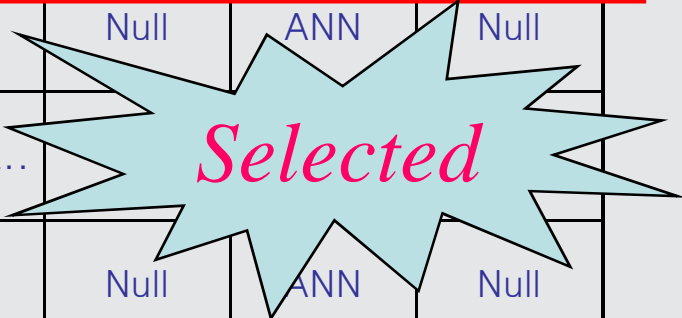
- No. Cat subCat Upnode Node Equiv
- 위 column으로 테이블을 잘라 $I(\text{workingTable})$ 계산
- $E(\text{upperNode})$, $E(\text{Category})$, $E(\text{subCategory})$ 계산
- $I - E = \text{Gain}$
 - UpperNode > Category > SubCategory 순으로 높은 값이 계산됨

6. 검색

■ 검색 알고리즘

- Node에 InformationTheory가 있으면 1순위 검색
- Node에 InformationTheory는 없지만 {Equivalent, one of, intersectionOf, unionOf, restrictionOf, onProperty}가 있으면 Node에 해당 속성에 해당하는 Term을 넣고 검색 (2순위)
 - intersectionOf ID3
 - ID3를 Information Theory가 있는 Node열에 넣음
- InformationTheory와 같은 upperNode를 갖는 데이터 검색 (3순위)
- Node에 ID3를 넣은 후 같은 upperNode를 갖는 데이터 검색 (4순위)

6. 검색

1_2	Course	DM	Theory	InfTh.	Null	Null	Null	ID3
2_1	PaperW	BackStd	Theory	InfTh.	Null	Null	Null	ID3
1_1	Course	DM	Algo.	ID3	Null	Null	Null	InfTh.
5_1	TA	AI(U)	Algo.	ID3	Null	Null	Null	InfTh.
3_1	Course	ML	Algo.	PCA	Null	Null	ANN	Null
3_2	Course	ML	Domain	ANN	Artificia..			
4_1	Course	ANN	Algo.	MLPBP	EBPA			Null
6_1	TA	DS(U)	Theory	Graph	Null	Null	Null	Search
6_2	TA	DS(U)	Algo.	Search	Null	Null	Null	Graph

7. 결론 및 토의

■ Lab Works

- 종이 Lab Note의 단점
 - 검색이 불편
 - 보관 및 정리가 어려움
- 온톨로지를 이용하여 연구업무 뿐 아니라 연구실 업무 전반에 관해 체계적으로 정리할 수 있도록 설계
- Information Theory를 이용
 - 검색 시 이전의 작업(혹은 연구) 중에서 원하는 정보와 가장 가까운 작업 리스트를 보여줄 수 있도록 함

7. 결론 및 토의

■ 문제점

- 그림이나 표 삽입이 불편

■ 고려되지 않은 사항 및 반성점

- Text mining분야에서 유사도 측정방법을 몇 개 찾을 수 있었으나 비교 분석을 하지 못했음
- 제시한 방법의 타당성 검증을 보다 신뢰성 있게 하지 못함
 - 샘플데이터의 숫자가 매우 적었음
- Topic Idea를 너무 늦게 잡아서 시간에 쫓김

감사합니다