



인 터 넷 진 화 의 열 쇠

온톨로지

웹 2.0에서 3.0으로

노상규·박진수 공저

god's toy business, 2007

SNU IDB laboratory

Ontology Contents

MODULE 1 온톨로지의 개념 및 응용

- Chapter 1 온톨로지 개요
 - 1. 온톨로지의 유래
 - 2. 분류와 개념화 과정
 - 3. 컴퓨터 온톨로지
- Chapter 2 온톨로지의 분류와 용도
 - 1. 온톨로지의 분류
 - 2. 온톨로지의 사용 목적과 중요성
 - 3. 온톨로지와 시맨틱 웹
- Chapter 3 온톨로지 구축 프로젝트
 - 1. 사이크(Cyc)
 - 2. 워드넷(WordNet)
 - 3. 전자거래문서
 - 4. 통합의학언어시스템
 - 5. 오픈 디렉터리 프로젝트
 - 6. 국제상품분류코드(UNSPSC)
- Chapter 4 온톨로지 적용 분야
 - 1. 전자상거래 분야
 - 2. 의료 분야
 - 3. 법률 분야
 - 4. 검색 서비스 분야
 - 5. 문화컨텐츠 분야

MODULE 2 온톨로지 언어와 구축도구

- Chapter 5 온톨로지 언어
 - 1. 온톨로지 언어의 발전 과정
 - 2. 인공지능 기반의 온톨로지 언어
 - 3. 온톨로지 마크업 언어
- Chapter 6 RDF(S): RDF와 RDF Schema
 - 1. XML과 RDF
 - 2. RDF
 - 3. RDF Schema
 - 4. RDF(S)의 한계점
- Chapter 7 OWL(Web Ontology Language)
 - 1. OWL의 기본 요소: 클래스와 속성
 - 2. OWL의 새로운 기능
 - 3. 세 종류의 OWL
 - 4. OWL 예제
- Chapter 8 토픽맵(Topic Maps)과 XTM(XML Topic Maps)
 - 1. 토픽맵(Topic Maps) 개념
 - 2. 토픽맵 구성요소
 - 3. XTM 예제
- Chapter 9 온톨로지 틀
 - 1. 온톨로지 틀의 분류
 - 2. 온톨로지 개발 틀
 - 3. 주요 온톨로지 틀 요약 정보

Chapter 3 온톨로지 구축 프로젝트

■ 특정 목적을 가지는 시스템의 핵심 구성 요소로서 온톨로지를 구축한 사례

1. 지식추출 : 인간의 지식 + 추론 엔진 → 인간과 같은 사고를 하는 컴퓨터



2. 자연어 검색 : 일상적으로 흔히 사용되는 구문 정보



3. 분야별 어휘 사전 : 특수 분야에서 사용하는 용어 용어들 간의 관계



4. 분류체계 용도 : 디렉터리, 상품 분류



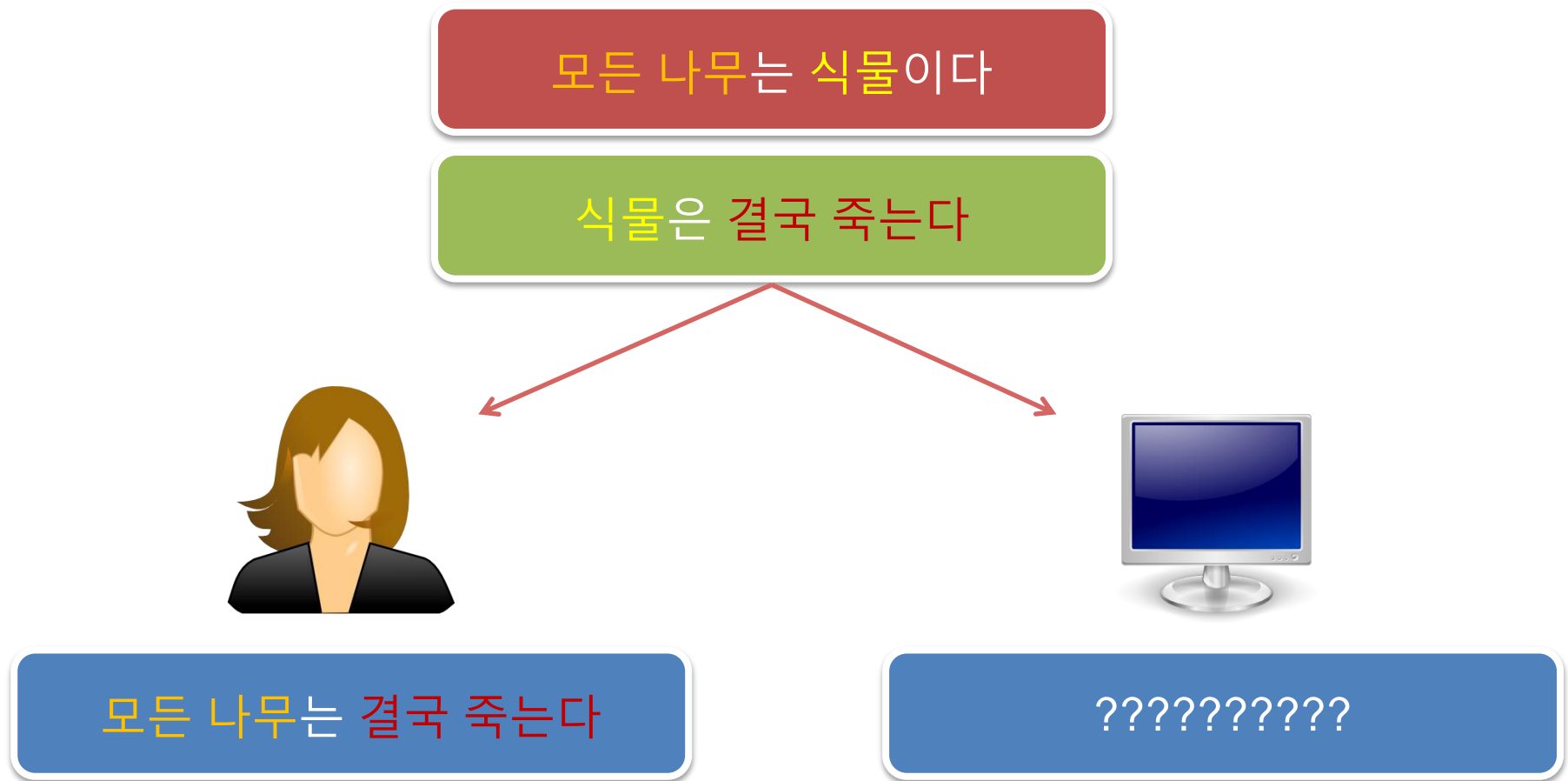
unspsc

Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)

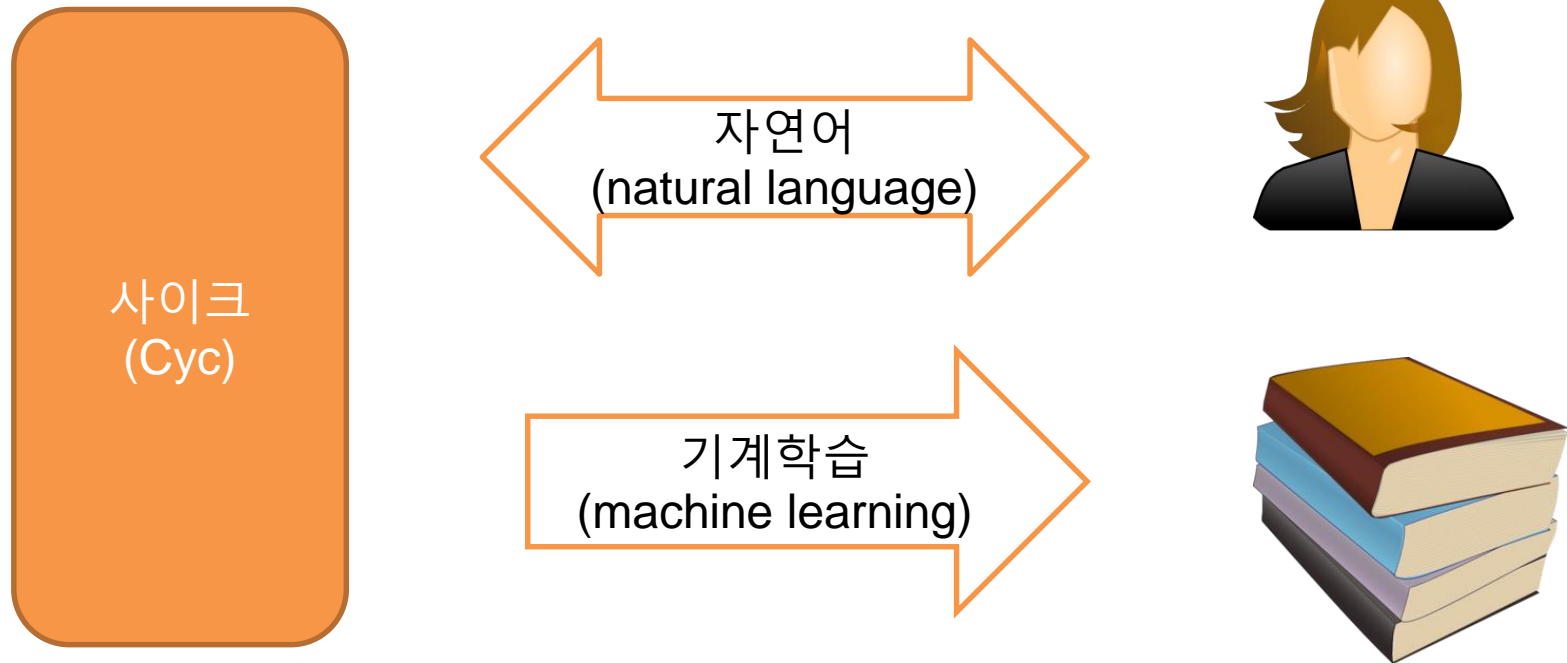
1. 사이크(Cyc)

- 컴퓨터가 인간과 비슷한 지식을 가질 수 있도록 도와주는 역할

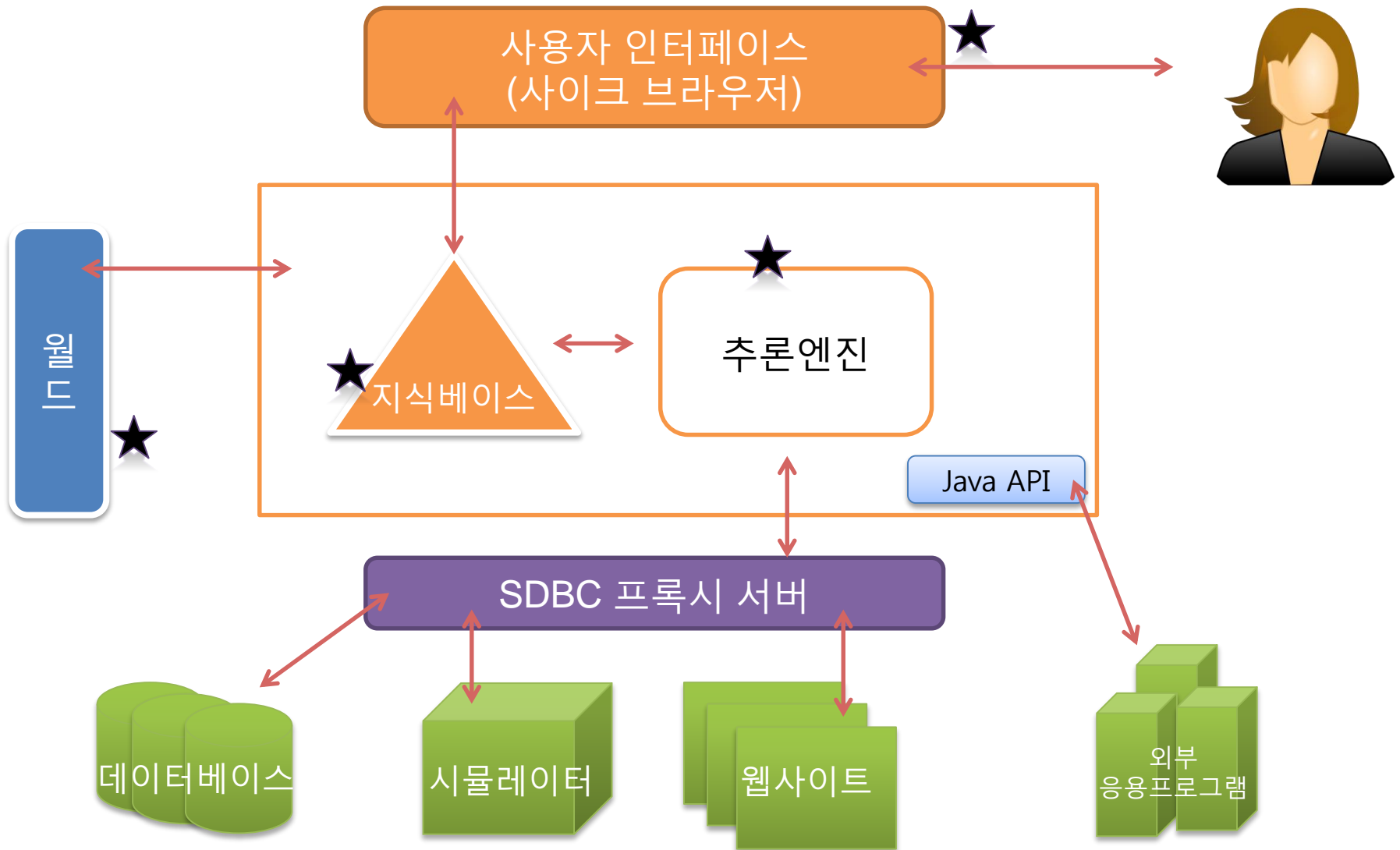


1. 사이크(Cyc)

- 인간이 가지고 있는 일반상식과 전문 지식을 모두 포함하는 지식베이스를 구축



1. 사이크(Cyc) : 구조



1. 사이크(Cyc) : 구성요소 (1/2)

■ 지식베이스 (knowledge base)

- 2백만 개 이상의 선언적 사실 (assertion) + 15만개 이상의 개념 (concept)
- 인간의 지식 보유 + 컴퓨터가 지식을 추출 → 사이크의 온톨로지
- 어휘와 문법적인 내용 포함 → 사람과 컴퓨터가 자연어를 통해 의사소통

■ 추론 엔진 (inference engine)

- 지식베이스에 저장되어 있는 선언적 사실 + 외부에서 얻는 정보
→ 결론에 도달하거나 새로운 추측
- 한번에 여러 가지 문제를 동시에 해결 가능

1. 사이크(Cyc) : 구성요소 (2/2)

■ 월드 (World)

- 이전에 사용했던 선언적 사실들을 스냅샷으로 저장
- 새로운 사이크가 수행될 때, 이전의 내용들은 무결성 확인 없이 즉시 메모리로 로드

■ 인터페이스 (user interface)

- 사용자가 지식베이스의 내용을 질의하거나 탐색, 수정하는 브라우저

1. 사이크(Cyc) : 이용분야

- 미 국방부의 해커 탐지용 시스템
- 인터넷 검색엔진 라이코스(Lycos)



- 제약회사 – 글락소스미스클라인 (GlaxoSmithKline)

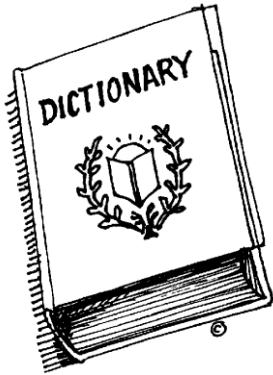
Cyc: 현황

- OpenCyc Ontology: 2001년 Cyc Ontology의 일부가 웹사이트를 통해 공개
- OpenCyc can be used as the basis of a wide variety of intelligent applications such as:
 - rich domain modeling
 - semantic data integration
 - text understanding
 - domain-specific expert systems
 - game AIs
- ResearchCyc: 연구용 온톨로지 라이선스 배포중
- EnterpriseCyc: 기업들의 intelligent solution을 위한 basis를 제공

Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)

2. 워드넷 (WordNet)



WordNet

- 단어 중심으로 작성
- 단어를 통한 접근

- 문맥에서의 의미를 중심으로 작성
- 개념을 통한 접근

- 의미를 중심으로 접근할 수 있는 일반 사전의 기능
 - + 동의어, 반의어를 제시할 수 있는 시소러스 (thesaurus)
- 자동 텍스트 분석과 인공지능을 지원

2. 워드넷 (WordNet)

- 단어들을 Synset이라고 불리는 인지적 동의어들의 집합으로 분류
- Synset들을 의미관계로 엮어 놓음

- 하의어 (Hyponym)
- 상위어 (Hypernym)
- 동의어 (Sister term)

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **article** (nonfictional prose forming an independent part of a publication)
- [S:](#) (n) **article** (one of a class of artifacts) "*an article of clothing*"
 - [direct hyponym](#) / [full hyponym](#)
 - [S:](#) (n) [article of commerce](#) (an article that is offered for sale)
 - [S:](#) (n) [breakable](#) (an article that is fragile and easily broken) "*pack the breakables separately*"
 - [S:](#) (n) [knickknack](#), [novelty](#) (a small inexpensive mass-produced article)
 - [S:](#) (n) [notion](#) ((usually plural) small personal articles or clothing or sewing items) "*buttons and needles are notions*"
 - [S:](#) (n) [ware](#) (articles of the same kind or material; usually used in combination: 'silverware', 'software')
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
- [S:](#) (n) **article**, [clause](#) (a separate section of a legal document (as a statute or contract or will))
- [S:](#) (n) **article** ((grammar) a determiner that may indicate the specificity of reference of a noun phrase)

Verb

- [S:](#) (v) **article** (bind by a contract; especially for a training period)

워드넷: 현황

- Created and Maintained by Princeton University
- WordNet has been used for a number of different purposes in information systems, including
 - word sense disambiguation
 - information retrieval, automatic text classification
 - automatic text summarization
 - machine translation
 - and even automatic crossword puzzle generation.

POS	Unique Strings	Synsets	Total Word-Sense Pairs
Noun	117,798	82,115	146,312
Verb	11,529	13,767	25,047
Adjective	21,479	18,156	30,002
Adverb	4,481	3,621	5,580
Totals	155,287	117,659	206,941

Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)

3. 전자거래문서

- 전자문서 : 컴퓨터 등 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성, 송수신 또는 저장된 문서



ROSETTANET

- UN의 전자거래, 무역촉진포럼 (UN/CEFACT)과 민간표준화기구인 OASIS가 추진하고 있는 XML 기반의 국제 표준
- 범용성이 높음

- 기업간 전자상거래 표준
- 기업이 직접 적용하고 있기 때문에 실용적

상호연동

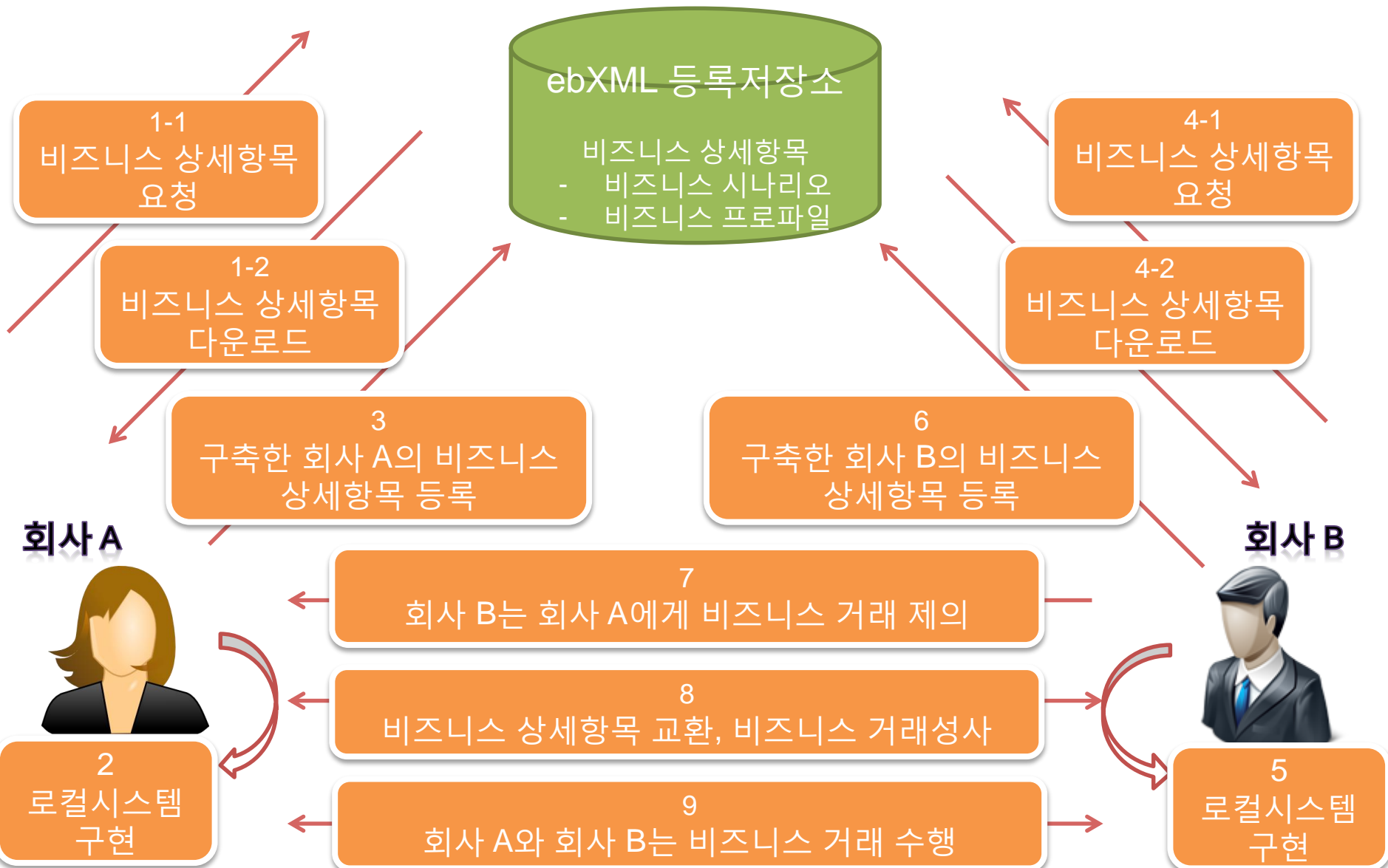


3.1 ebXML : 구성요소

- 전세계적으로 단일한 XML 기반의 개방형 전자상거래 표준



3.1 ebXML : 개념도



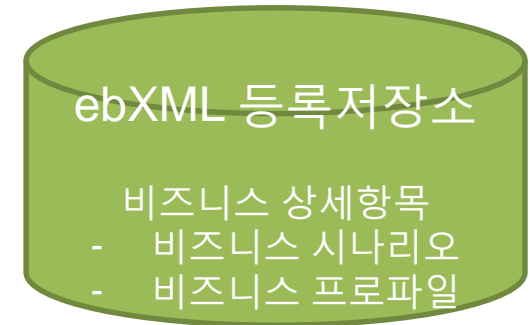
3.1 ebXML : 장, 단점

■ 장점

- ebXML 등록저장소의 표준에 맞추기만 하면 무료로 전자문서 교환 가능

■ 단점

- 아직까지는 일부 산업에만 적용



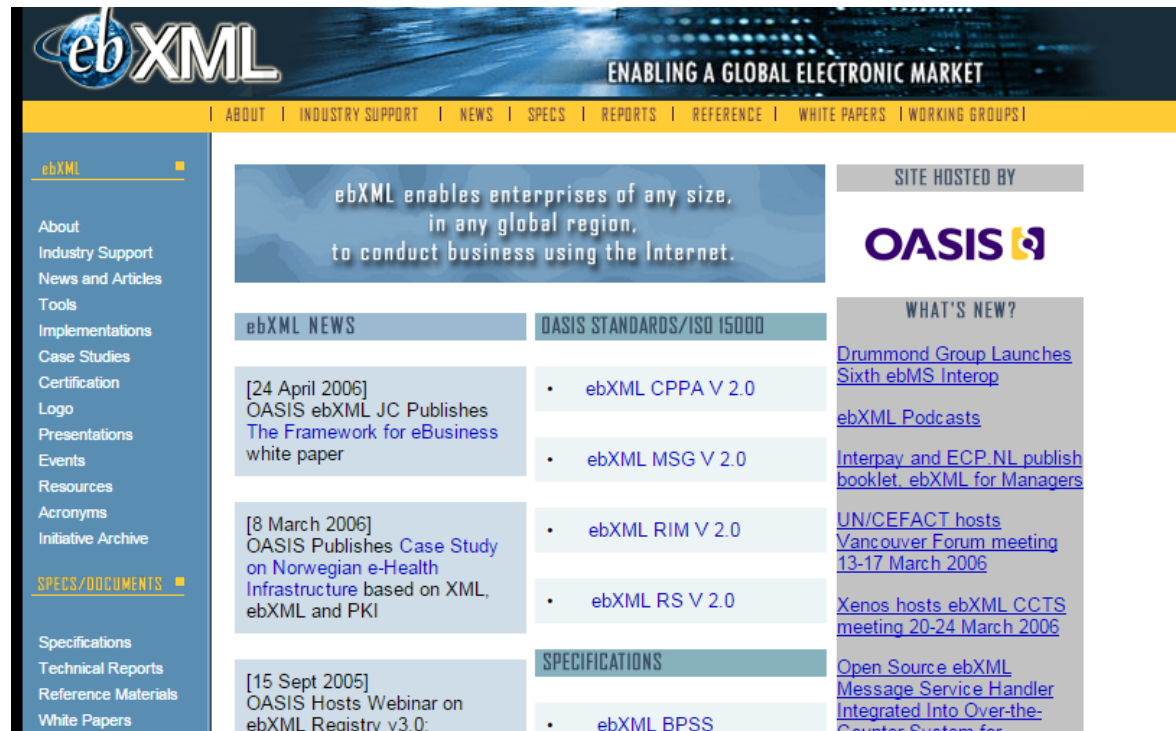
■ 이용분야

posco

SAMSUNG
삼성카드

ebXML: 현황

- Started in 1999
- Maintained by OASIS (민간표준화기구)
 - Advancing open standards for the information society



3.2 로제타넷 (RosettaNet)

- 정보기술, 전자부품, 반도체 제조 분야에 종사하는 전 세계 500개 이상의 기업체들이 제안한 전자거래문서 표준
- 개방형 표준 제공 : 비즈니스 사전 + 기술 사전 + 구현 프레임워크 + 비즈니스 메시지 관리체계 + 프로세스 정의

용어사전

- 비즈니스 사전(RNBD) / 기술사전(RNTD)
- 여러 회사에서 다양하게 사용되는 용어들을 정의

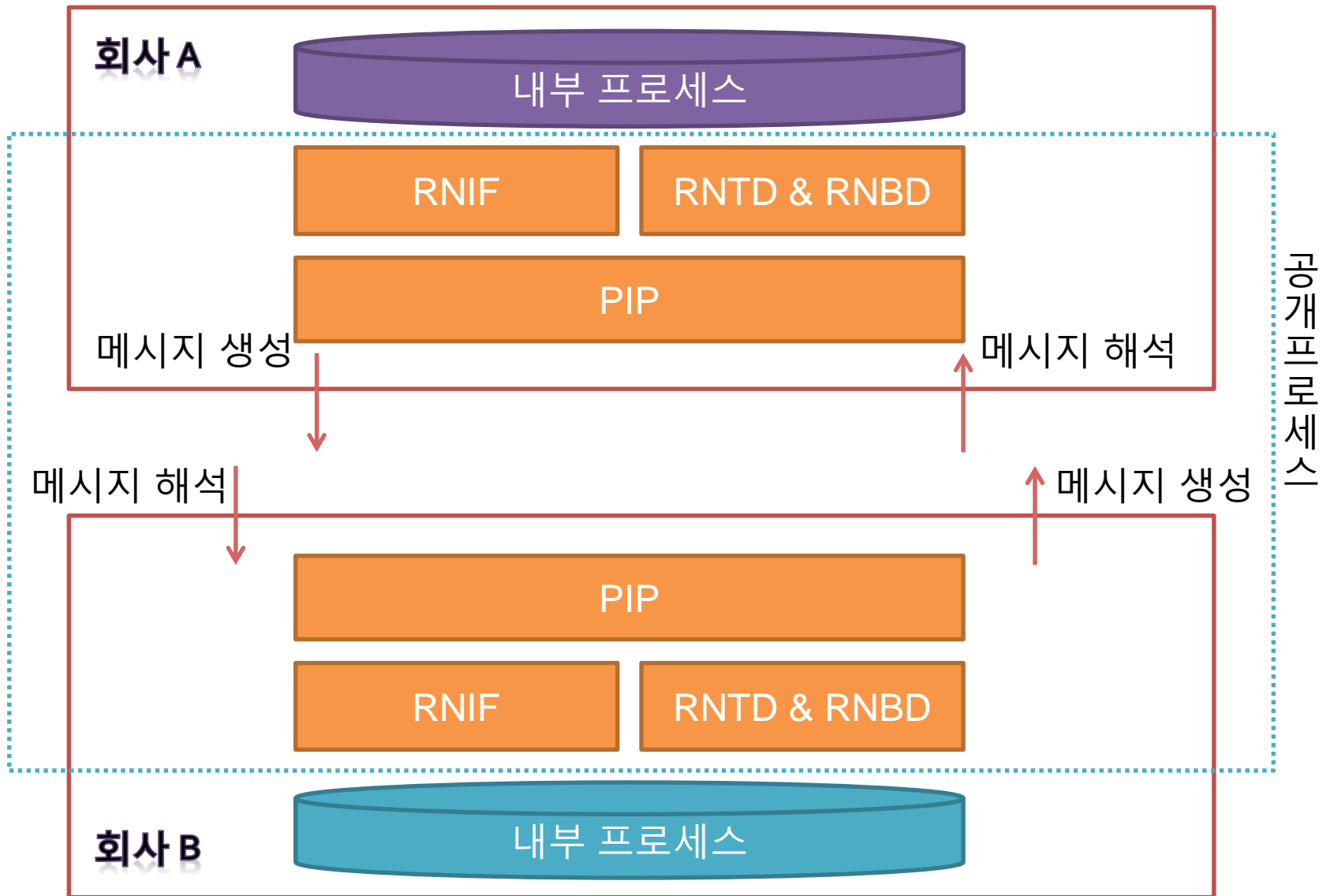
RNIF

- RosettaNet Implementation Framework
- 개발 가이드라인 + 통신 프로토콜 + 보안 방법

PIP

- 비즈니스 프로세스를 XML 문서로 정의
- 비즈니스 프로세스의 최소단위

3.2 로제타넷 (RosettaNet) : 개념도



3.2 로제타넷 (RosettaNet) : 장, 단점

■ 장점

- 기업의 필요에 의해서 생성되었기 때문에 현실적인 검증을 마침

■ 단점

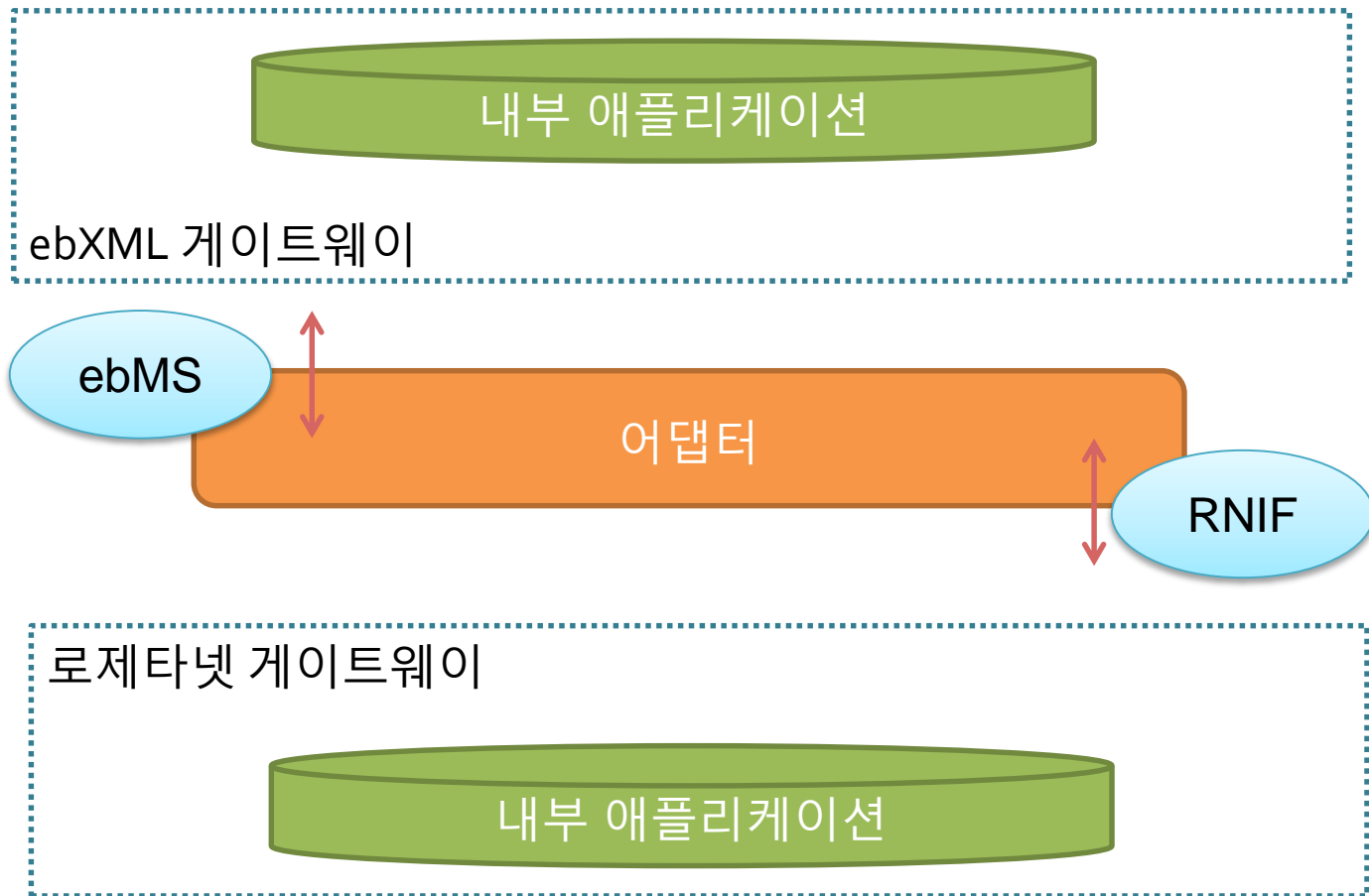
- 향후에 로열티 문제가 발생할 가능성
- 여러 분야의 산업에 적용 불가능

■ 이용분야



3.3 ebXML과 로제타넷의 표준화 방안

- 각자의 독립된 구조를 유지
- 상호 연동을 가능하게 하는 어댑터를 사용하여 교류



Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)

4. 통합의학언어시스템 (Unified Medical Language System)

- 의학 및 의료분야에서 사용되는 표준화된 용어체계
- 각기 다른 시스템과 데이터베이스간에 호환성을 보장

메타소러스

- 개념 또는 의미로 구성
- 동의어/ 유사어/ 변형형 등을 그룹화

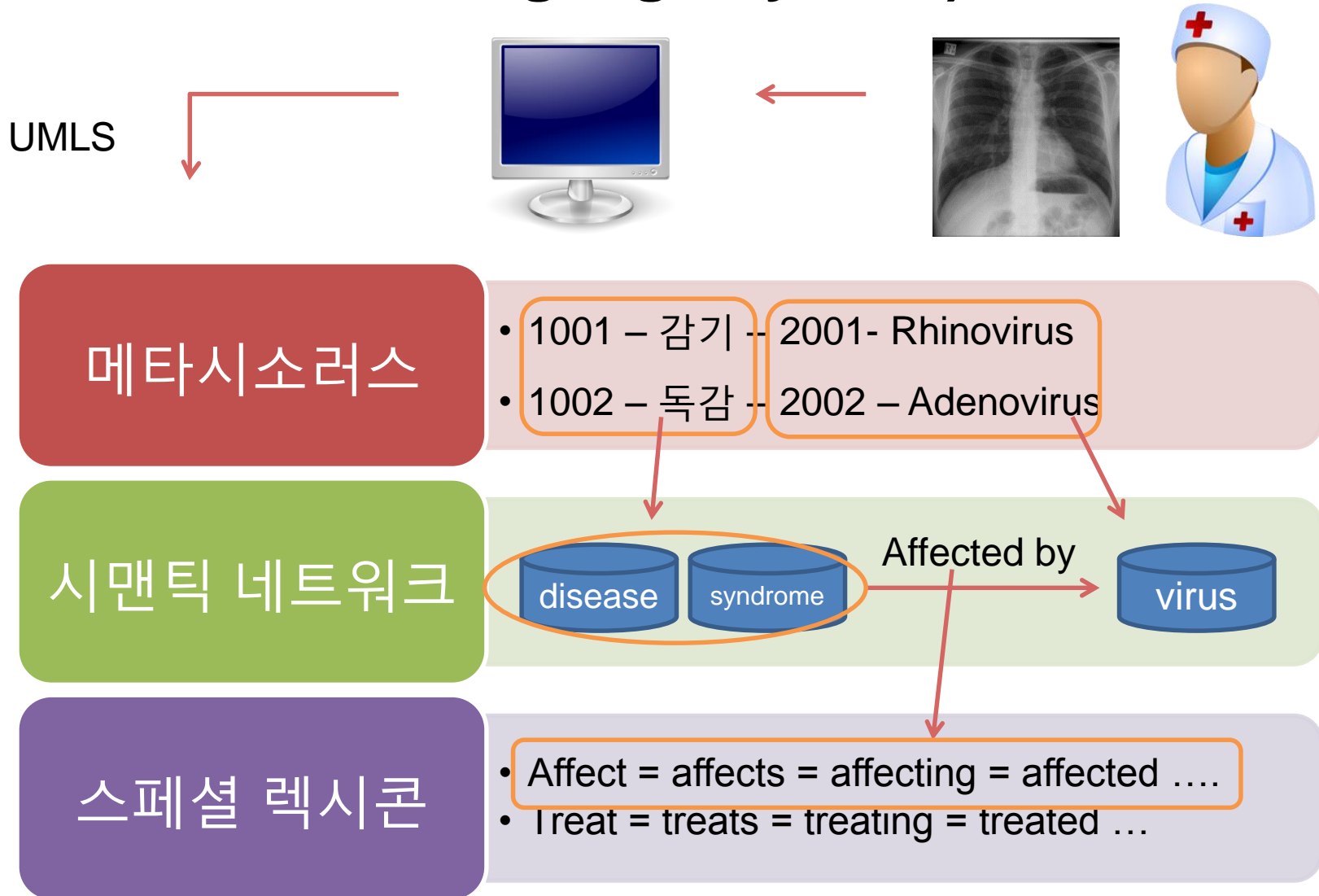
시맨틱 네트워크

- 노드와 링크로 이루어진 지식표현 스키마
- 메타소러스의 개념을 분류하고 개념간의 관계를 나타냄

스페셜 렉시콘

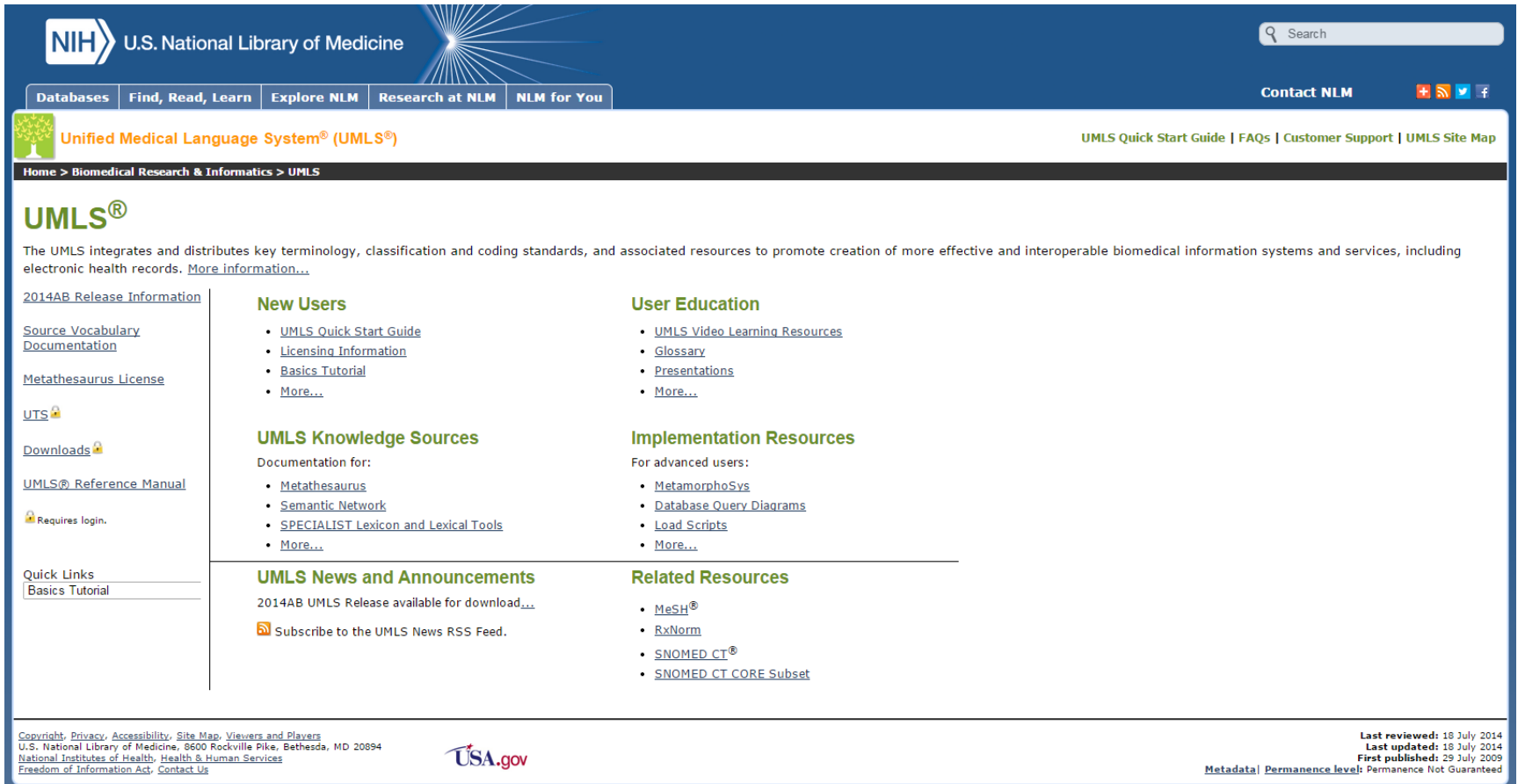
- 메타소러스에 포함되지는 않지만 필요한 용어와 자연어 처리를 위한 구문정보

4. 통합의학언어시스템 : 구조 (Unified Medical Language System)



Unified Medical Language System: 현황

- Created in 1986
- Maintained by US National Library of Medicine, NIH



The screenshot shows the UMLS website homepage. At the top is the NIH logo and 'U.S. National Library of Medicine'. Below this is a navigation bar with links: Databases, Find, Read, Learn, Explore NLM, Research at NLM, NLM for You. A search bar is on the right. The main header area includes the UMLS logo and a description: 'The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records.' Below this, there are several sections: '2014AB Release Information', 'Source Vocabulary Documentation', 'Metathesaurus License', 'UTS', 'Downloads', 'UMLS® Reference Manual', 'Quick Links', 'Basics Tutorial', 'New Users', 'User Education', 'UMLS Knowledge Sources', 'Implementation Resources', 'UMLS News and Announcements', and 'Related Resources'. The footer contains copyright information, contact details, and a 'Last reviewed' date of 18 July 2014.

NIH U.S. National Library of Medicine

Search

Databases Find, Read, Learn Explore NLM Research at NLM NLM for You

Contact NLM

Unified Medical Language System® (UMLS®)

UMLS Quick Start Guide | FAQs | Customer Support | UMLS Site Map

Home > Biomedical Research & Informatics > UMLS

UMLS®

The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records. [More information...](#)

[2014AB Release Information](#)

[Source Vocabulary Documentation](#)

[Metathesaurus License](#)

[UTS](#)

[Downloads](#)

[UMLS® Reference Manual](#)

[Requires login.](#)

Quick Links

[Basics Tutorial](#)

New Users

- [UMLS Quick Start Guide](#)
- [Licensing Information](#)
- [Basics Tutorial](#)
- [More...](#)

User Education

- [UMLS Video Learning Resources](#)
- [Glossary](#)
- [Presentations](#)
- [More...](#)

UMLS Knowledge Sources

Documentation for:

- [Metathesaurus](#)
- [Semantic Network](#)
- [SPECIALIST Lexicon and Lexical Tools](#)
- [More...](#)

Implementation Resources

For advanced users:

- [MetamorphoSys](#)
- [Database Query Diagrams](#)
- [Load Scripts](#)
- [More...](#)

UMLS News and Announcements

2014AB UMLS Release available for download...

[Subscribe to the UMLS News RSS Feed.](#)

Related Resources

- [MeSH®](#)
- [RxNorm](#)
- [SNOMED CT®](#)
- [SNOMED CT CORE Subset](#)

Copyright, Privacy, Accessibility, Site Map, Viewers and Players
U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health, Health & Human Services
Freedom of Information Act, Contact Us

USA.gov

Last reviewed: 18 July 2014
Last updated: 18 July 2014
First published: 29 July 2009
Metadata | Permanence level: Permanence Not Guaranteed

Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)


5. 오픈 디렉터리 프로젝트 (Open Directory Project)

- 자발적인 인터넷 사용자가 참여하여 만든 WWW의 공개 디렉터리
- 웹 콘텐츠 분류체계용 온톨로지를 구축하는 것이 목표
- 디렉터리 간의 관계를 계층적으로 담고 있다
- 디렉터리 서비스를 제공하는 이유
 - 넓은 카테고리로부터 어떻게 검색을 좁혀나가야 하는지 모를 때 유용하게 쓰임
 - 한 영역 내의 토픽들이 어떻게 연관 되어 있는지를 이해하는데 도움을 줌
 - 검색에 유용한 용어들을 제안함

5. 오픈 디렉터리 프로젝트 : 예제 (Open Directory Project)

■ 서울에 위치하고 있는 대학교의 리스트

- Reference → Education → Colleges and Universities → Asia → South Korea → Seoul

 open directory project

In partnership with
Aol Search.

[about dmoz](#) | [dmoz blog](#) | [suggest URL](#) | [update listing](#) | [become an editor](#) | [report abuse/spam](#) | [help](#)

the entire directory ▼

[Top](#): [Regional](#): [Asia](#): [South Korea](#): [Seoul](#): [Education](#): Colleges and Universities (0) [Description](#)

- [Berea University of Graduate Studies](#)@ (1)
- [Capital Baptist Theological Seminary](#)@ (0)
- [Chongshin University](#)@ (0)
- [Chugye University for the Arts](#)@ (1)
- [Dankook University](#)@ (1)
- [Dongguk University](#)@ (1)
- [Duksung Women's University](#)@ (1)
- [Ewha Womans University](#)@ (4)
- [Hansung Technical College](#)@ (0)
- [Hansung University](#)@ (1)
- [Hanyang University](#)@ (1)
- [Honam University](#)@ (1)
- [Hongik University](#)@ (1)
- [International Design School for Advanced Studies](#)@ (1)
- [International Graduate School of English](#)@ (0)
- [KDI School of Public Policy and Management](#)@ (1)
- [Konkuk University](#)@ (1)
- [Kookmin University](#)@ (0)
- [Kyung Hee University](#)@ (2)
- [Methodist Theological Seminary](#)@ (1)
- [Myongji University](#)@ (0)
- [Namseoul University](#)@ (0)
- [Presbyterian College and Theological Seminary](#)@ (1)
- [Sahmyook University](#)@ (0)
- [Sangmyung University](#)@ (1)
- [Sejong University](#)@ (0)
- [Seokyeong University](#)@ (0)
- [Seoul National University](#)@ (0)
- [Seoul National University of Education](#)@ (0)
- [Seoul National University of Technology](#)@ (0)
- [Seoul Sports Graduate University](#)@ (0)
- [Seoul Women's University](#)@ (1)
- [Seoul-Jeongsu Polytechnic College](#)@ (0)
- [Sogang University](#)@ (1)
- [Sookmyung Women's University](#)@ (1)
- [Soongsil University](#)@ (0)

오픈 디렉터리 프로젝트: 현황

- https://www.dmoz.org/Computers/Internet/Searching/Directories/DMOZ/Sites_Using_DMOZ_Data/
- AOL이 소유하고 있지만 자발적인 편집자 모임이 유지 및 관리중
- 오픈 디렉터리를 사용하고 있는 사이트
 - ABC.net
 - Delphi Search Engine Directory
 - 등등 20여 개 사이트

Chapter 3 온톨로지 구축 프로젝트

- 1. 사이크 (Cyc)
- 2. 워드넷 (WordNet)
- 3. 전자거래문서
- 4. 통합의학언어시스템
- 5. 오픈 디렉터리 프로젝트
- 6. 국제상품분류코드 (UNSPSC)

6. 국제상품분류코드 (UNSPSC)

- 전 세계적으로 가장 널리 알려지고 활용되고 있는 전 산업대상의 전자상거래용 상품분류체계



- 상품을 명확하게 식별하는 코드
- 다른 상품과의 관련성을 표현하지 못함
- 효율적으로 상품 검색, 관리하기 위한 방법
- 상품을 계층관계로 정의한 단순한 형태의 온톨로지
- 8자리 숫자 값으로 표현

6. 국제상품분류코드 (UNSPSC) : 구성요소

세그먼트(Segment)

- 분석을 목적으로 하는 패밀리의 논리적인 집합 (aggregation)

패밀리(Family)

- 일반적으로 인정되는 상호관련성 있는 상품 분류

클래스(Class)

- 사용 혹은 기능을 공유하는 상품 그룹

커머디티(Commodity)

- 대체 가능한 상품 혹은 서비스의 그룹

6. 국제상품분류코드 (UNSPSC) : 예제

세그먼트	44				사무기기 & 용품
		10			사무용 기기
		11			사무용 가구
패밀리		12			사무용품
			17		필기구
			18		수정도구
클래스			19		잉크 및 펜/샤프심 리필
				01	잉크리필
				02	샤프심 리필
커미디티				03	펜심 리필



‘펜심 리필’ = UNSPSC Classification 44 – 12 – 19 - 03

6. 국제상품분류코드 (UNSPSC) : 이점



UNSPSC



- 표준화된 상품의 계층적인 분류체계
→ 전자상거래의 활성화

- 기업내 중복구매 방지
- 상호 비교분석을 위한 통계 제공

- 정확한 상품 검색 & 구매

국제상품분류코드 (UNSPSC): 현황

- Managed by GS1 US of the UN Development Programme (UNDP)
 - A neutral, not-for-profit, international organization that develops and maintains standards
- The latest released February 2014.
- Contains over 50,000 commodities.

기타

■ 비즈니스분야

- EO (Enterprise Ontology)
- TOVE (Toronto Virtual Enterprise)

■ 화학분야

- Chemical Ontology (Fernandez-Lopez et al, 1999)

■ 생물분야

- Gene Ontology

■ 약품분야

- DOPE (Drug Ontology Project)

■ 지리분야

- Spatial and Thematic Ontology