

FCAView Tab:

A Concept-Oriented View Generation Tool for Clinical Data Using Formal Concept Analysis

Guoqian Jiang, Katsuhiko Ogasawara, Naoki Nishimoto,

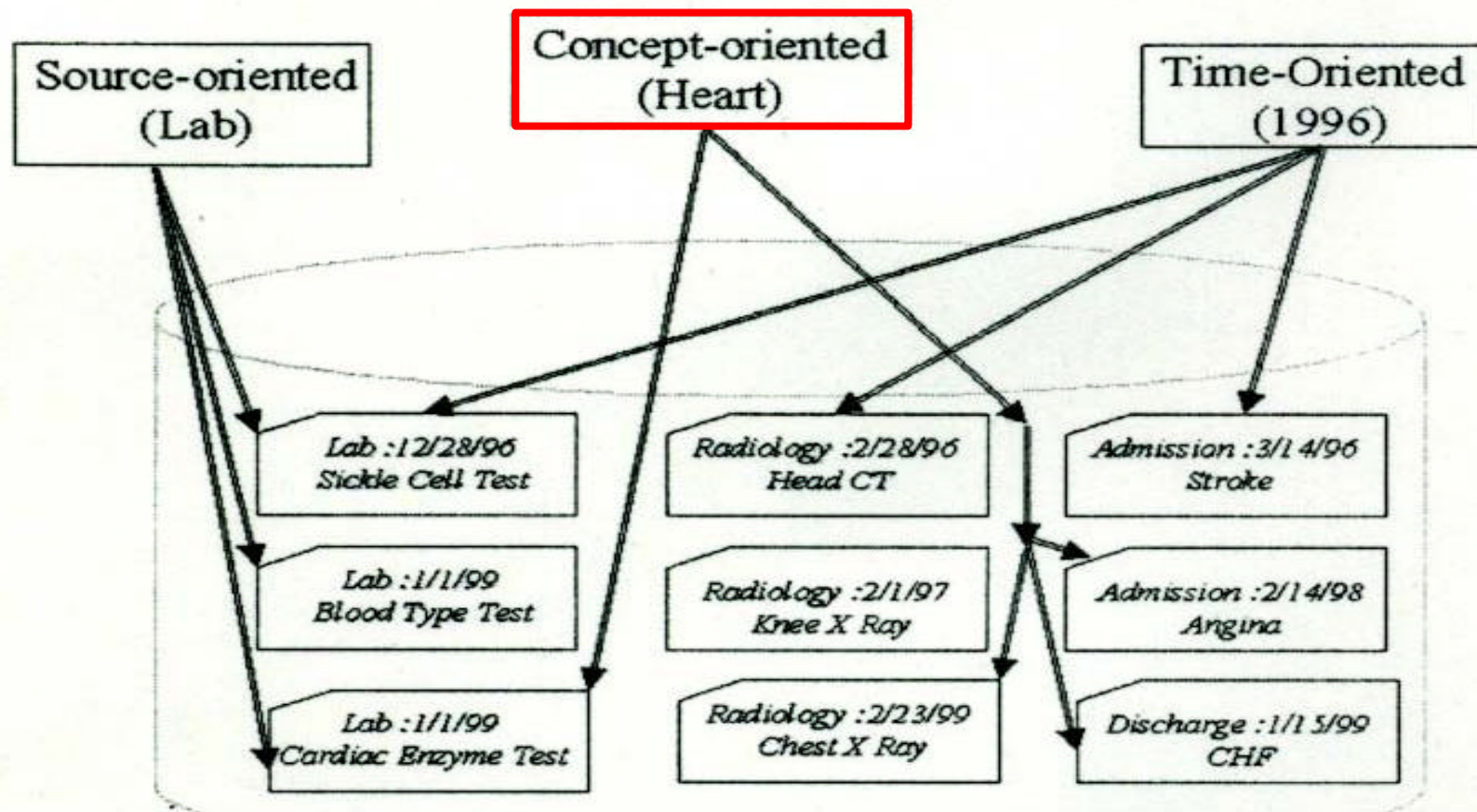
Akira Endoh, Tsunetaro Sakurai

Department of Medical Informatics, Hokkaido University
Graduate School of Medicine, Sapporo, Japan

The 8th International Protégé Conference

Madrid, Spain, July 20, 2005

Classification of Views for Electronic Patient Records



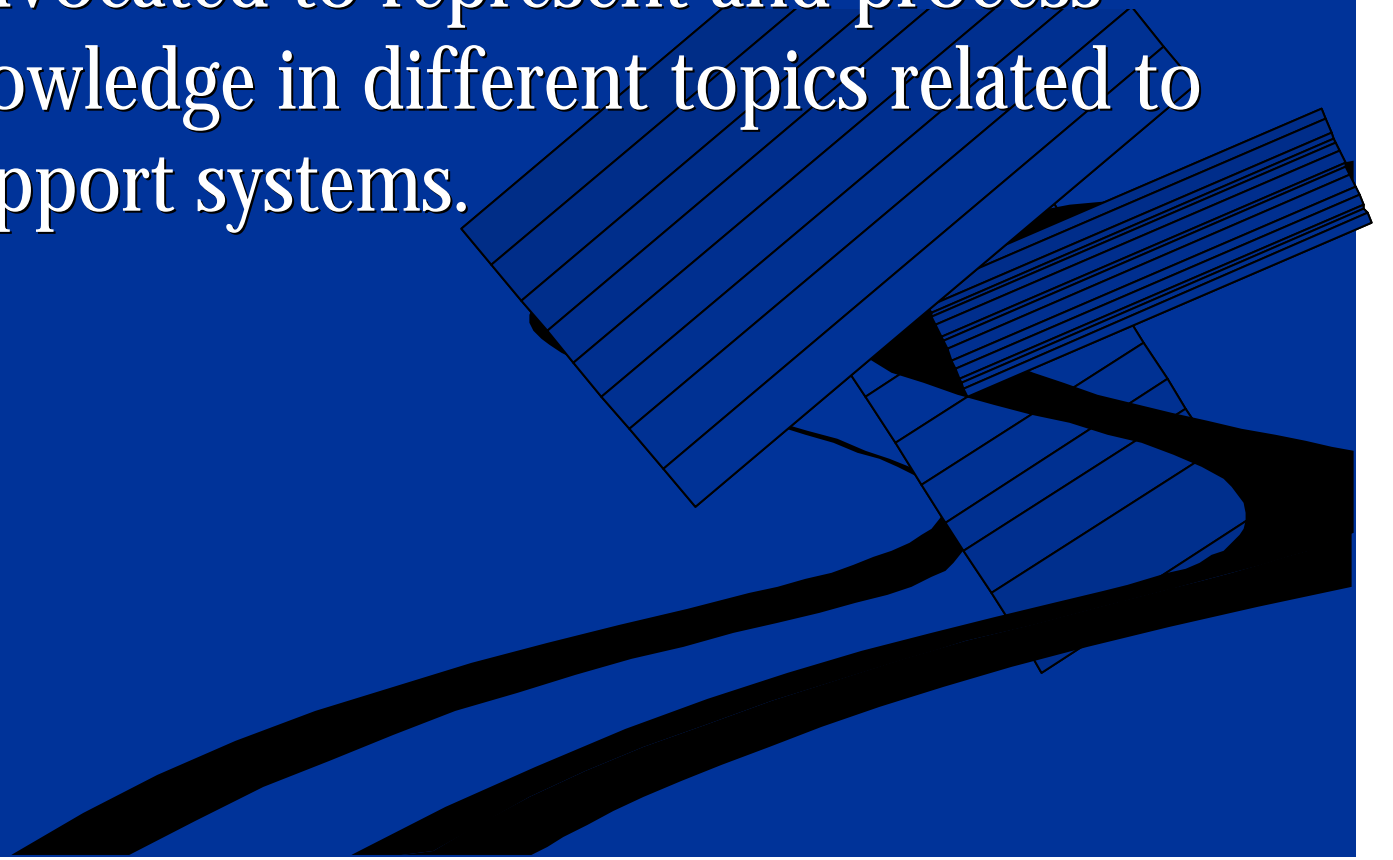
From Q.Zeng, et al. Journal of Biomedical Informatics. 2001;34(2):112-128

Concept-Oriented Views

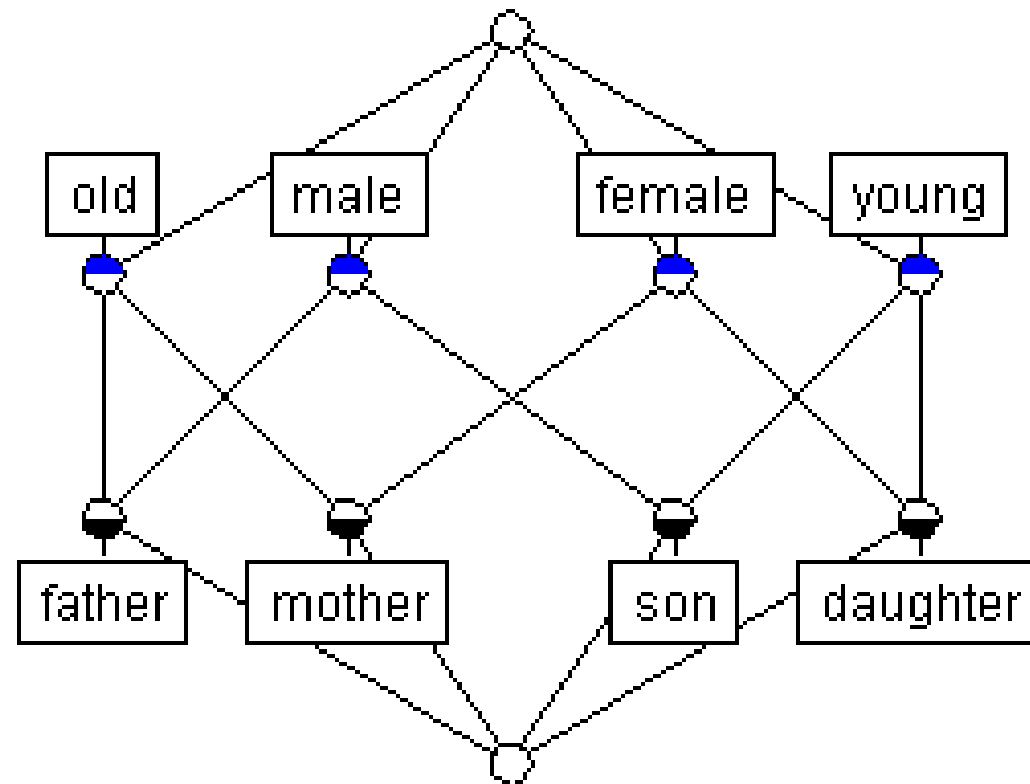
- ✍ Several studies have demonstrated that computer-generated, concept-oriented views for clinical data can be used to reduce clinicians' information overload and improve the accuracy of clinical data retrieval.
- ✍ However, the tasks for automating the generation of concept-oriented views are challenging.

Formal Concept Analysis

- ✍ Formal concept analysis (FCA), a mathematical approach to data analysis based on lattice theory, has been advocated to represent and process medical knowledge in different topics related to decision support systems.

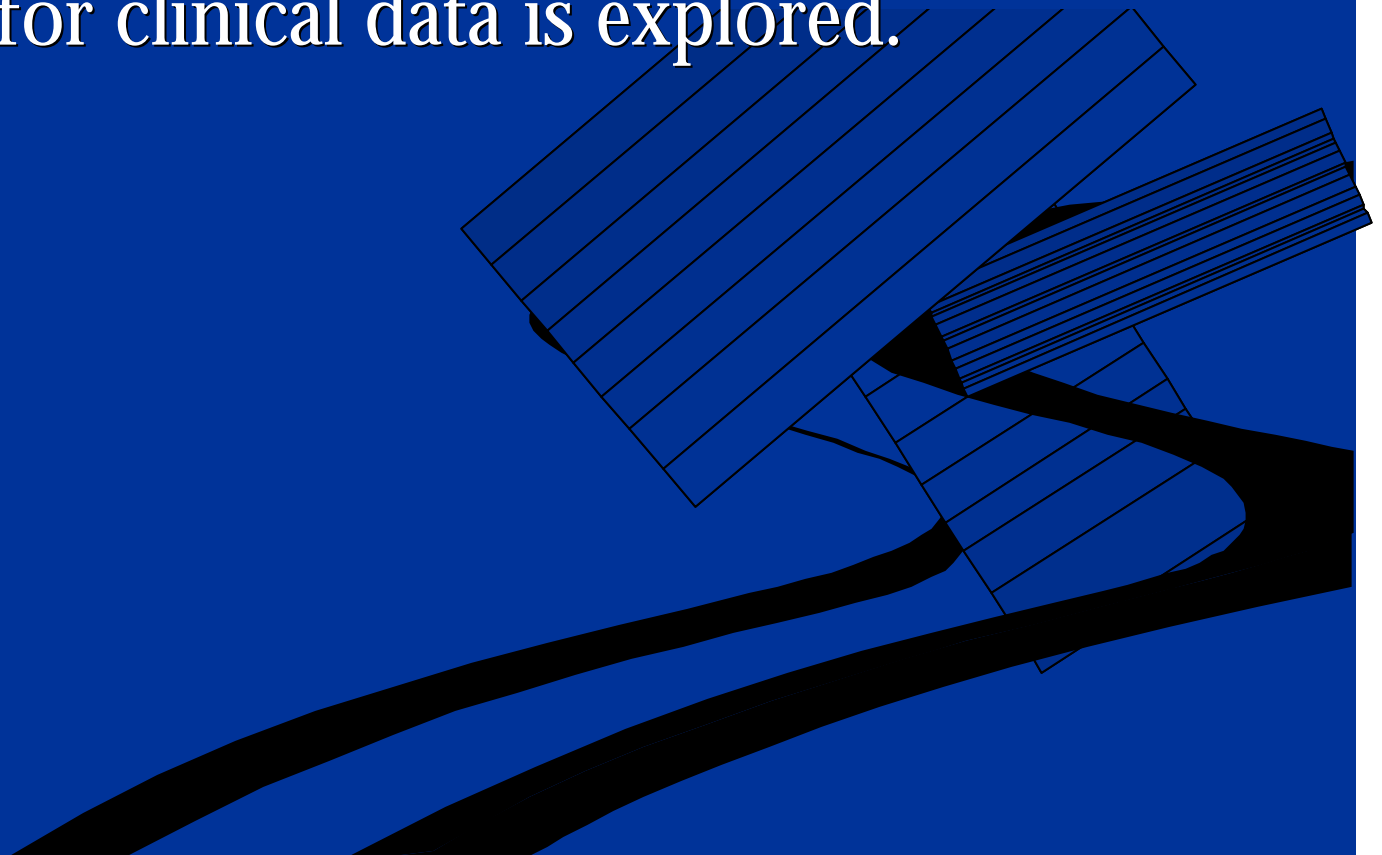


Example: A family



Study Purpose

- ✍ In the study, the potential role of formal concept analysis on concept-oriented view generation for clinical data is explored.



Materials

✍ Clinical Data

- ✍ A collection of 386 Japanese discharge summaries obtained from cardiovascular domain.

✍ Knowledge sources

- ✍ The Japanese version of MeSH (MSHJPN) in UMLS2005AA (54,550 entries).
 - ✍ Developed by the NLM
- ✍ MEDIS version 2.3 --The Japanese standard dictionaries of diagnosis terms (19,660 entries) and procedural terms (9,686 entries).
 - ✍ Developed by the MEDIS-DC in Japan

System Construction

✍ **FCAView Tab:** A Protégé-2000 tab plug-in

✍ A virtual patient records system

✍ A ontology-based knowledge base

✍ Medical concepts

✍ Formal concept analysis

✍ A natural language processing (NLP) module

✍ A formal concept analysis (FCA) module

The structure of discharge summary represented by a XML file

```
<?xml version="1.0" encoding="Shift_JIS"?>
<!DOCTYPE DischargeSummary SYSTEM "DischargeSummary.dtd">
<DischargeSummary>
  <PatientDischargeSummary>
    <FileInformation>
      <FileID>pcdata</FileID>
      <PatientID>pcdata</PatientID>
    </FileInformation>
    <PatientInformation>
      <PatientAge>pcdata</PatientAge>
      <PatientSex>pcdata</PatientSex>
      <PatientOccupation>pcdata</PatientOccupation>
    </PatientInformation>
    <DateInformation>
      <AdmissionDate>pcdata</AdmissionDate>
      <DischargeDate>pcdata</DischargeDate>
    </DateInformation>
    <DiagnosisList>
      <PrimaryDiagnosis>pcdata</PrimaryDiagnosis>
      <SecondaryDiagnosis>pcdata</SecondaryDiagnosis>
    </DiagnosisList>
    <PatientSummary>
      <Complaint>pcdata</Complaint>
      <CurrentHistory>pcdata</CurrentHistory>
      <PastHistory>pcdata</PastHistory>
      <OtherHistory>pcdata</OtherHistory>
      <ExaminationRecords>pcdata</ExaminationRecords>
      <InvestigationRecords>pcdata</InvestigationRecords>
      <ProgressRecords>pcdata</ProgressRecords>
    </PatientSummary>
  </PatientDischargeSummary>
</DischargeSummary>
```

An Example of the Virtual Patient Records

0001 (Instance of PatientDischargeSummary, internal name is XMLDischargeSummary_Instance_6)

FileInformationSlot		PatientInformationSlot			DateInformationSlot	
FileID	PatientID	PatientAge	PatientOccupation	PatientSex	DischargeDate	AdmissionDate
0001	****					

DiagnosisListSlot (1 values)

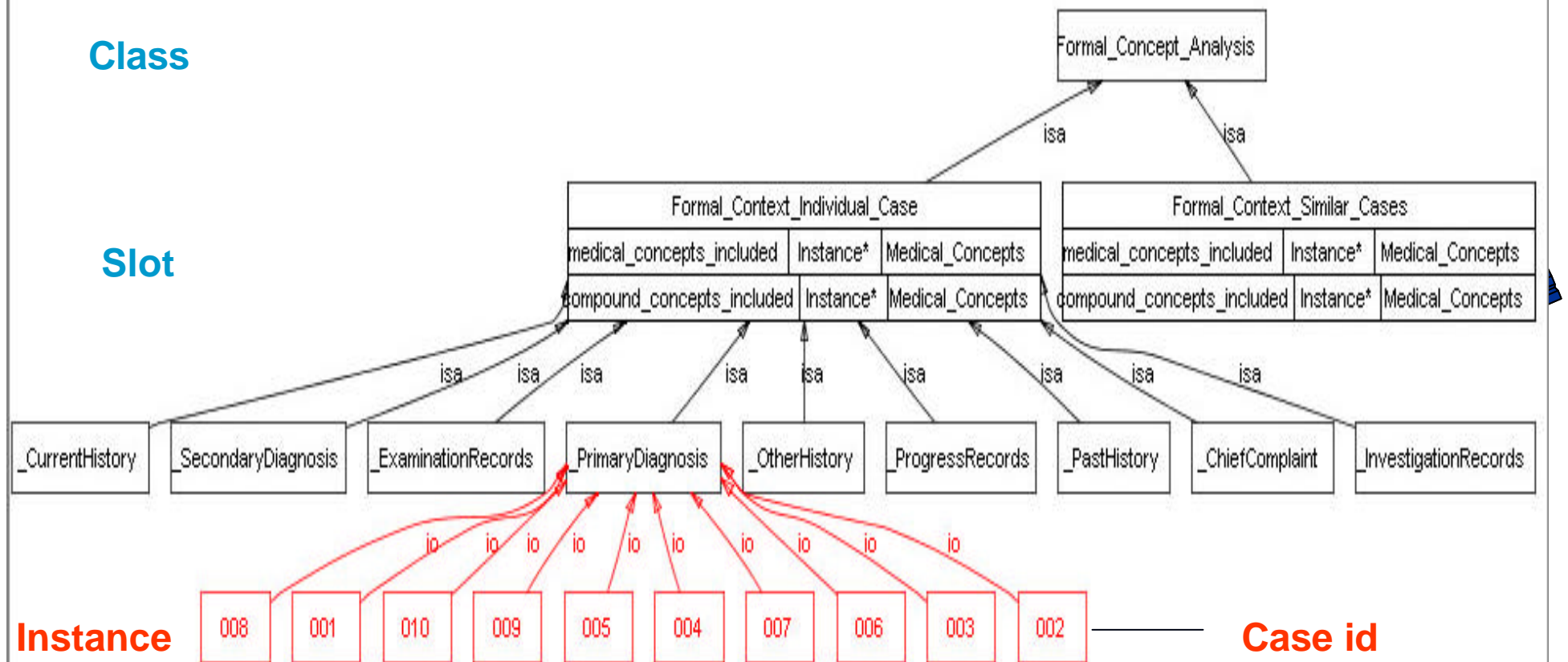
PrimaryDiagnosis

SecondaryDiagnosis

PatientSummarySlot (1 values)

Complaint	ExaminationRecords
CurrentHistory	InvestigationRecords
PastHistory	ProgressRecords
OtherHistory	

Ontology-based Knowledge Base



NLP Module

✍ ChaSen version 2.3

- ✍ A Japanese morphological analysis system
- ✍ Open source
- ✍ Developed by Nara Institute of Science and Technology

✍ MACD

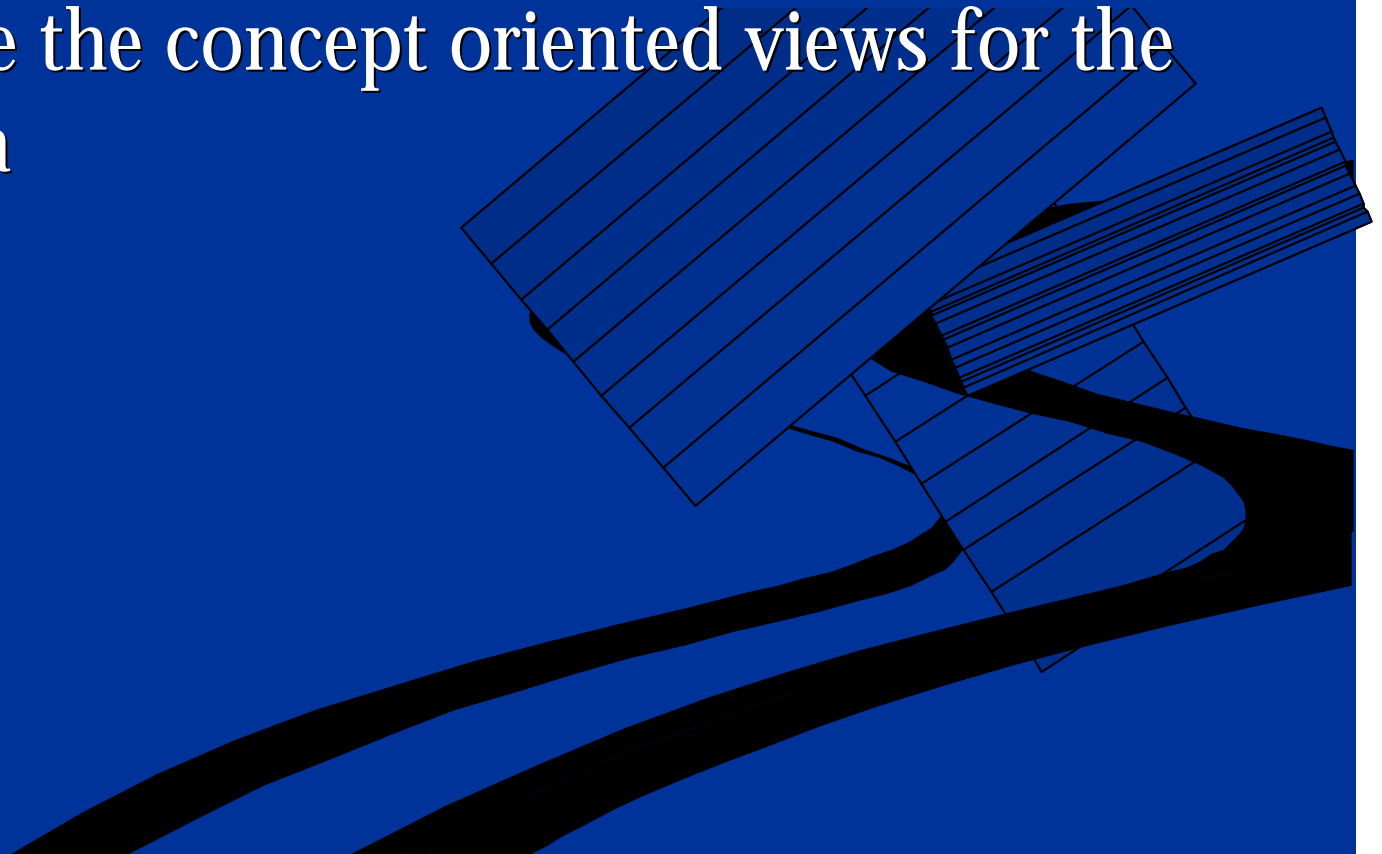
- ✍ Morphological Analyzer Connectivity Driver-model
- ✍ As the Java interface of ChaSen
- ✍ Developed by Nara Institute of Science and Technology

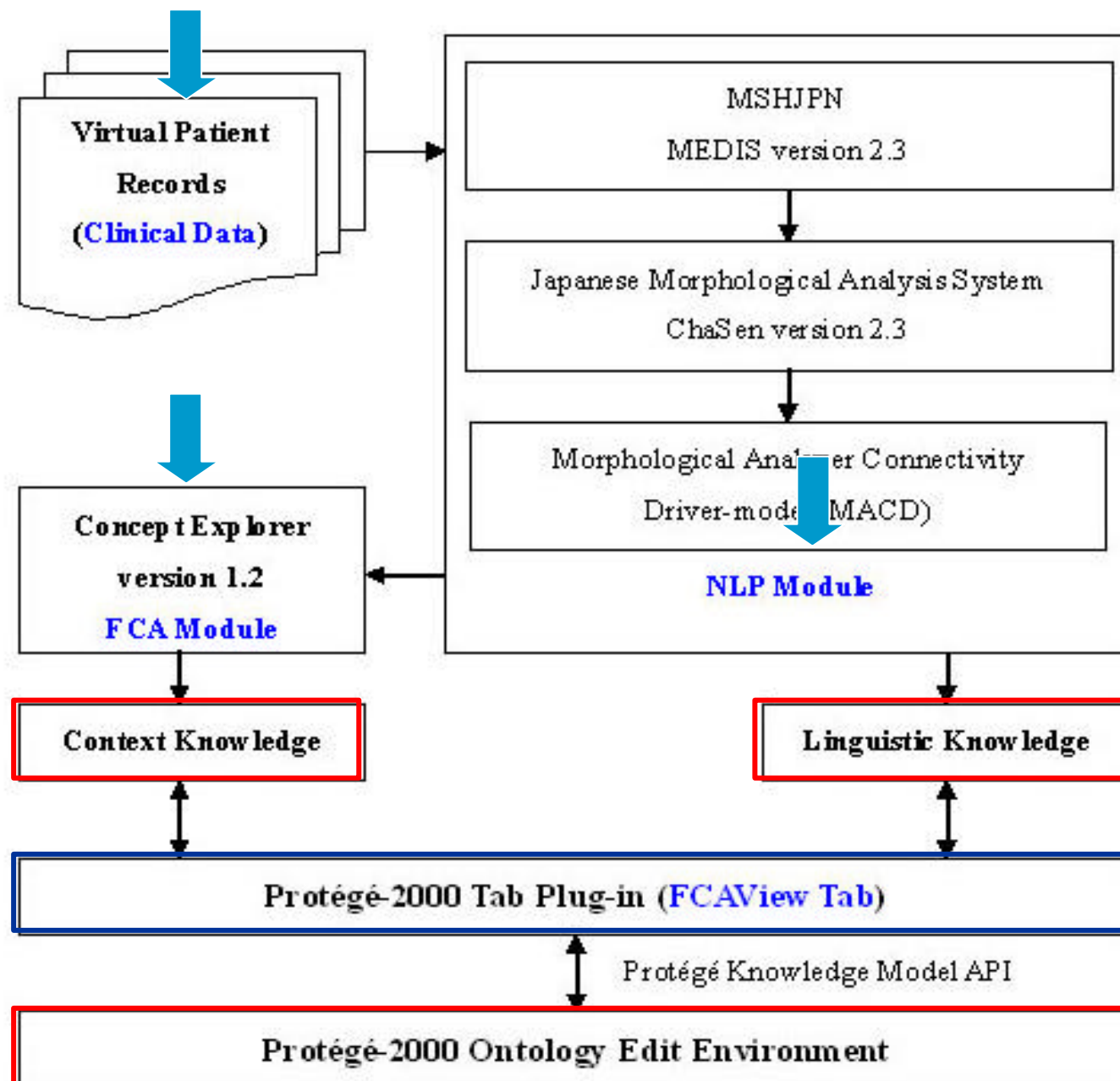
✍ User dictionaries:

- ✍ MSHJPN
- ✍ MEDIS version 2.3

FCA Module

- ✍ The Java API of Concept Explorer version 1.2
 - ✍ Open source project
- ✍ To generate the concept oriented views for the clinical data














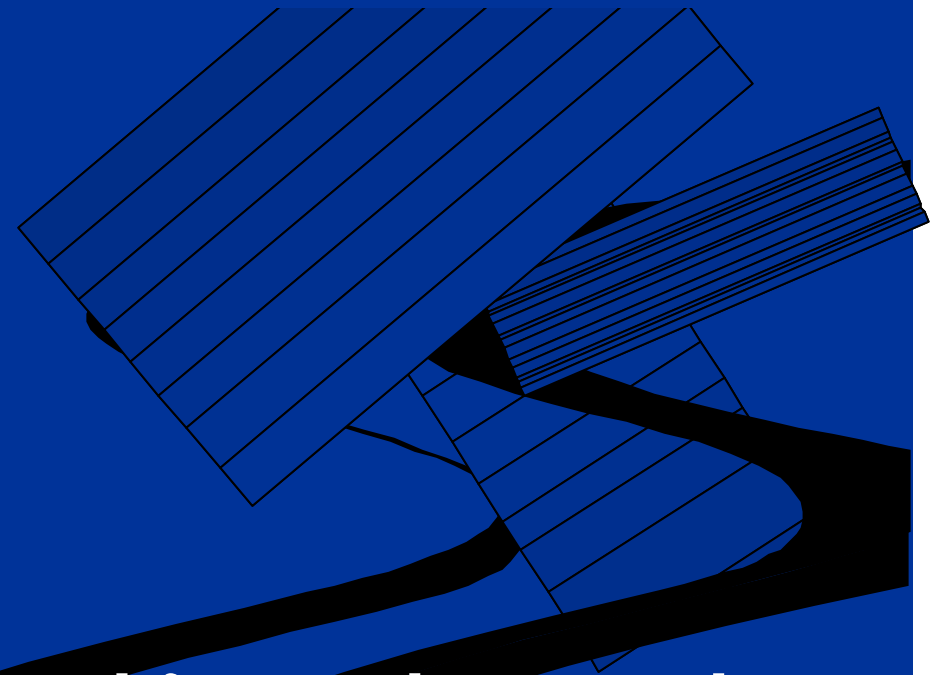
Results (I)

View generation for individual patient records

 The items about the structure knowledge of the patient records used as the formal objects

-  Primary diagnosis
-  Secondary diagnoses
-  Chief Complaints
-  Current history
-  Past history
-  Other histories
-  Examination records
-  Investigation records
-  Progress records

 The medical concepts extracted from each part used as the formal attributes



Context of an Individual Case by Cross-table

XMLDischargeSummary Protégé 3.0 (file:¥C:¥Projects¥Protege-projects¥FCA-dischargesummary¥XMLDischargeSummary.pprj, Default...

File Edit Project Window Help

Classes Queries FCAView Tab

Concept T (individual Case) NLP Module

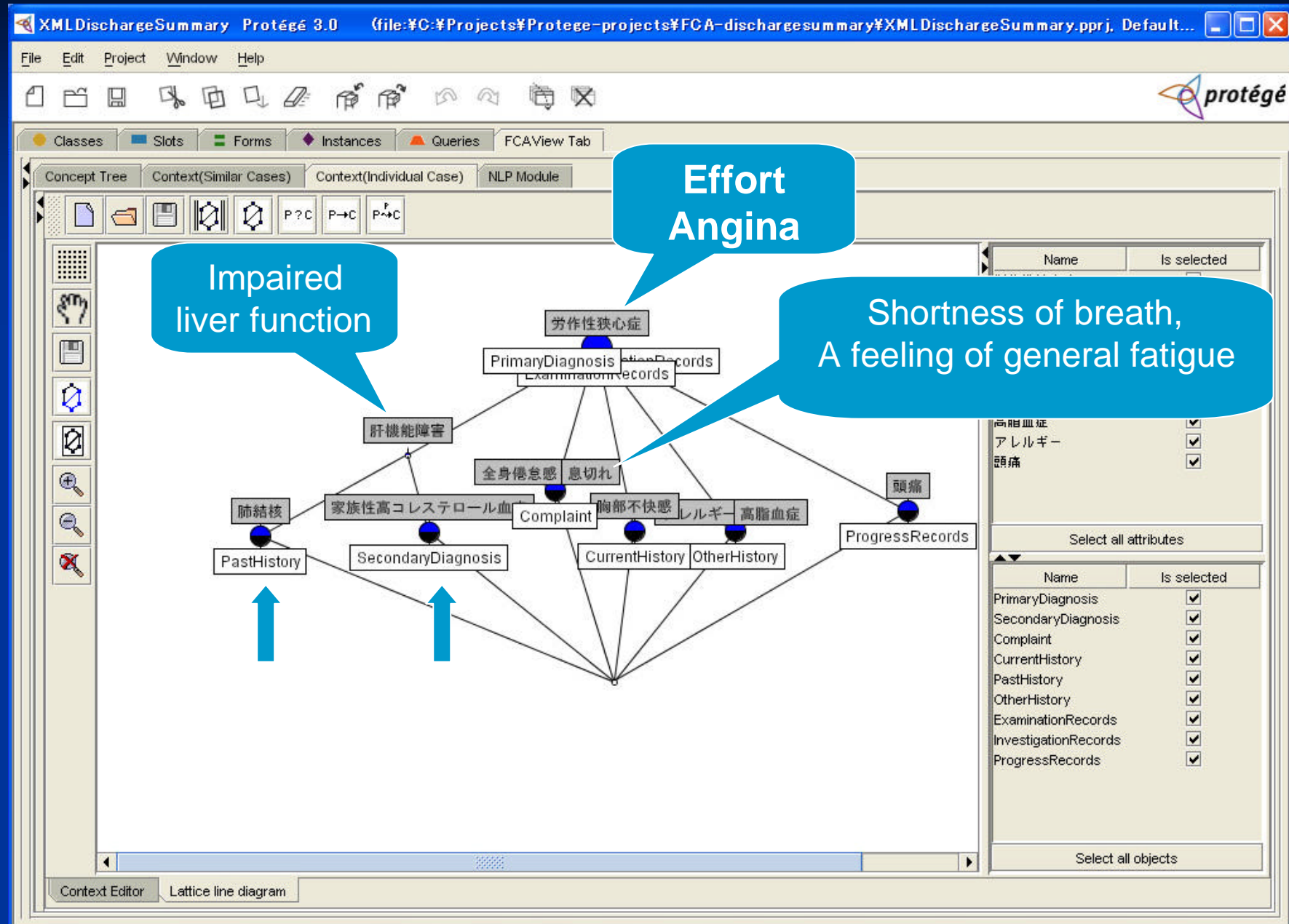
Structural Items

Medical Concepts

	B	C	D	E	F	G	H	I	J	K
	労作性狭心症	家族性高コ...	肝機能障害	息切れ	全身倦怠感	胸部不快感	肺結核	高脂血症	アレルギー	頭痛
PrimaryDiagnosis	X									
SecondaryDiagnosis	X	X	X		X					
Complaint	X			X	X					
CurrentHistory	X					X				
PastHistory	X		X				X			
OtherHistory	X							X	X	
ExaminationRecords	X									
InvestigationRecords	X									
ProgressRecords	X									X

Context Editor

Concept-Oriented View of an Individual Case by Lattice

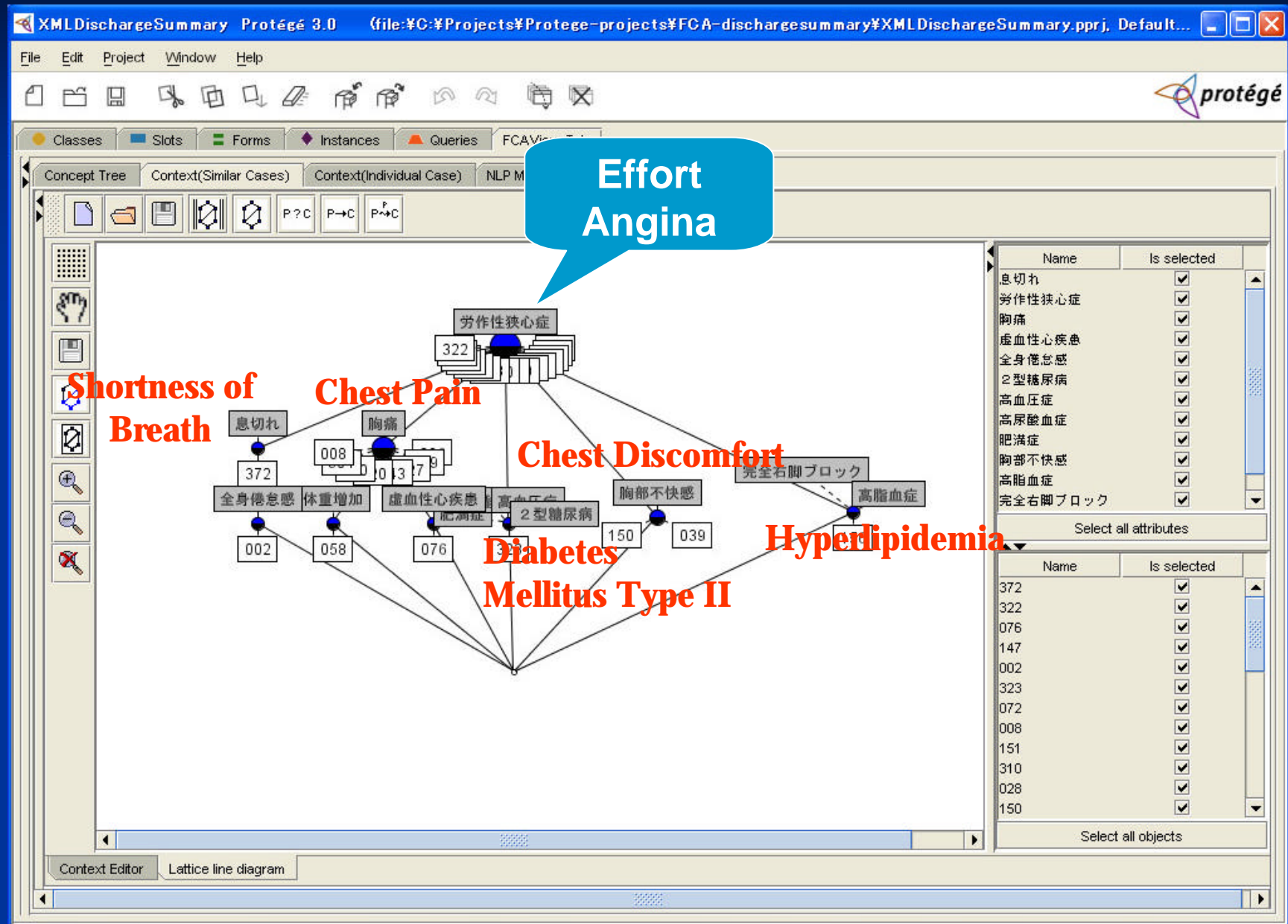


Results (II)

- ✍ View generation for the similar cases
 - ✍ The similar cases indicated by the case ids were used as the formal objects
 - ✍ The medical concepts extracted from each case were used as the formal attributes

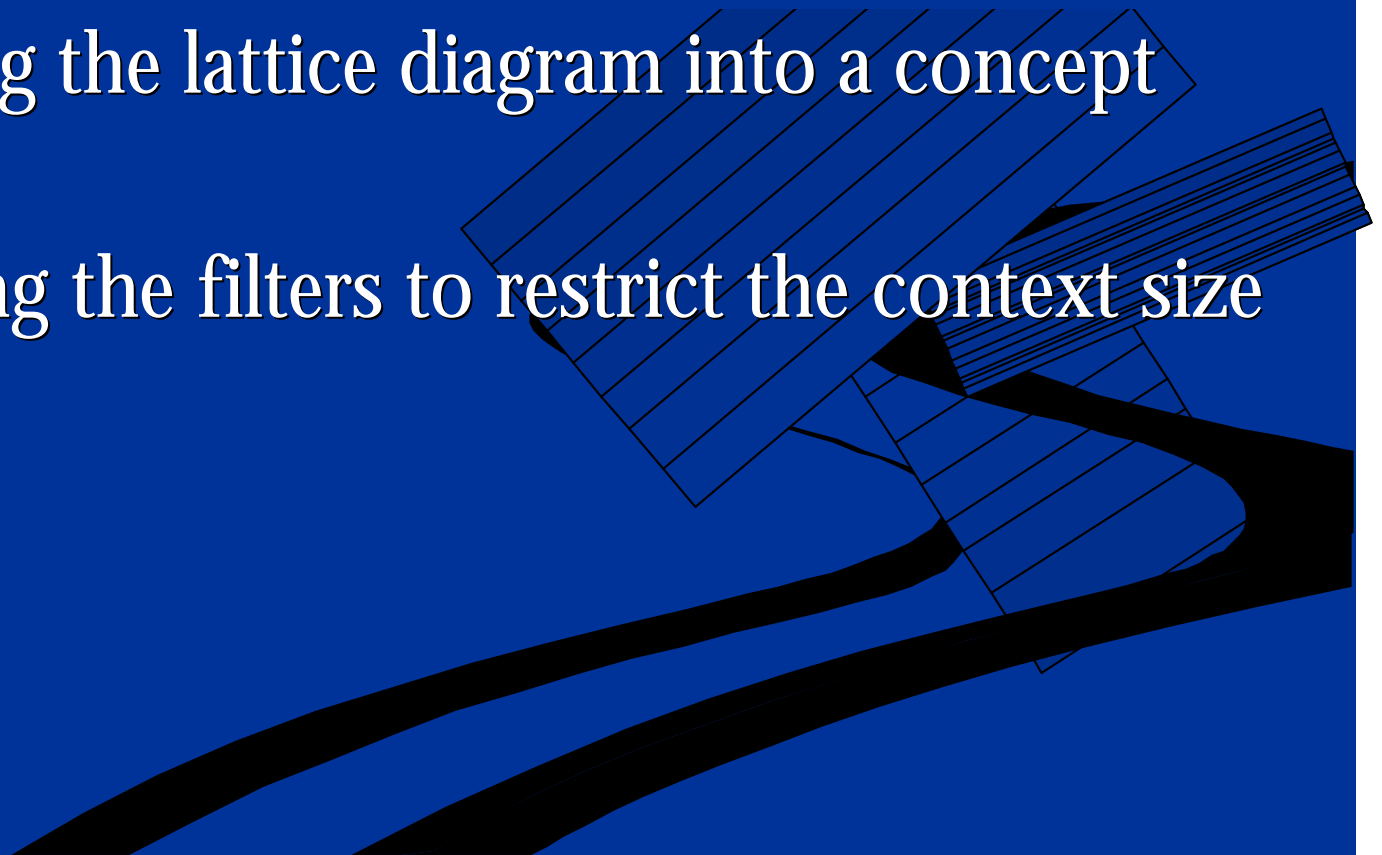
[illegible]

Concept-Oriented View of the Similar Cases by Lattice



Results (III)

- ✍ Implementation of information model for practical use
 - ✍ Converting the lattice diagram into a concept tree
 - ✍ Developing the filters to restrict the context size



XMLDischargeSummary Protégé 3.0 (file:¥C:¥Projects¥Protege-projects¥FCA-dischargesummary¥XMLDischargeSummary.pprj, Default...

File Edit Project Windows Help

For Context Of Effort Angina The Similar Cases

For Context Of The Individual Case Impaired Liver Function

Concept Tree: (368)

- 発作性上室性頻拍(2)
- 陳旧性心筋梗塞(13)
- 労作性狭心症(32)
- 心房細動(7)
- うっ血性心不全(12)
- 腹部大動脈瘤(1)
- 急性心筋梗塞(6)
- 肥大型心筋症(4)
- 拡張型心筋症(4)
- 胸腹部大動脈瘤(1)
- 急性上気道炎|感染性心内膜炎(1)
- 大動脈炎症候群(1)
- 収縮性心膜炎(2)
- 無症候性心筋虚血(6)
- 原発性アルドステロン症(2)
- 完全房室ブロック(4)
- 僧帽弁狭窄症(2)
- 失神(5)
- 胸膜炎(2)
- 深部静脈血栓症(1)
- 慢性腎不全(1)
- 肥大型非閉塞性心筋症(1)
- 脳幹梗塞(1)
- 洞不全症候群(3)
- 狭心症(9)
- リウマチ熱(1)
- 慢性心房細動(1)
- 脱水症|感冒(1)
- 大動脈弁閉鎖不全症(1)

Extent Instances:

- 002
- 008
- 028
- 035
- 039
- 043
- 058
- 059
- 060
- 070
- 072
- 076
- 120
- 147
- 150
- 151
- 156
- 217
- 227
- 228
- 251
- 270
- 277
- 286
- 304
- 310

Concept Tree: 002

- 労作性狭心症(9)
- 肝機能障害(2)
- 全身倦怠感|息切れ(1)
- 胸部不快感(1)
- アレルギー|高脂血症(1)
- 頭痛(1)

Structure Contents: 002

Basic Information

Age: Occupation: Sex: Admission Date: Discharge Date:

SecondaryDiagnosis

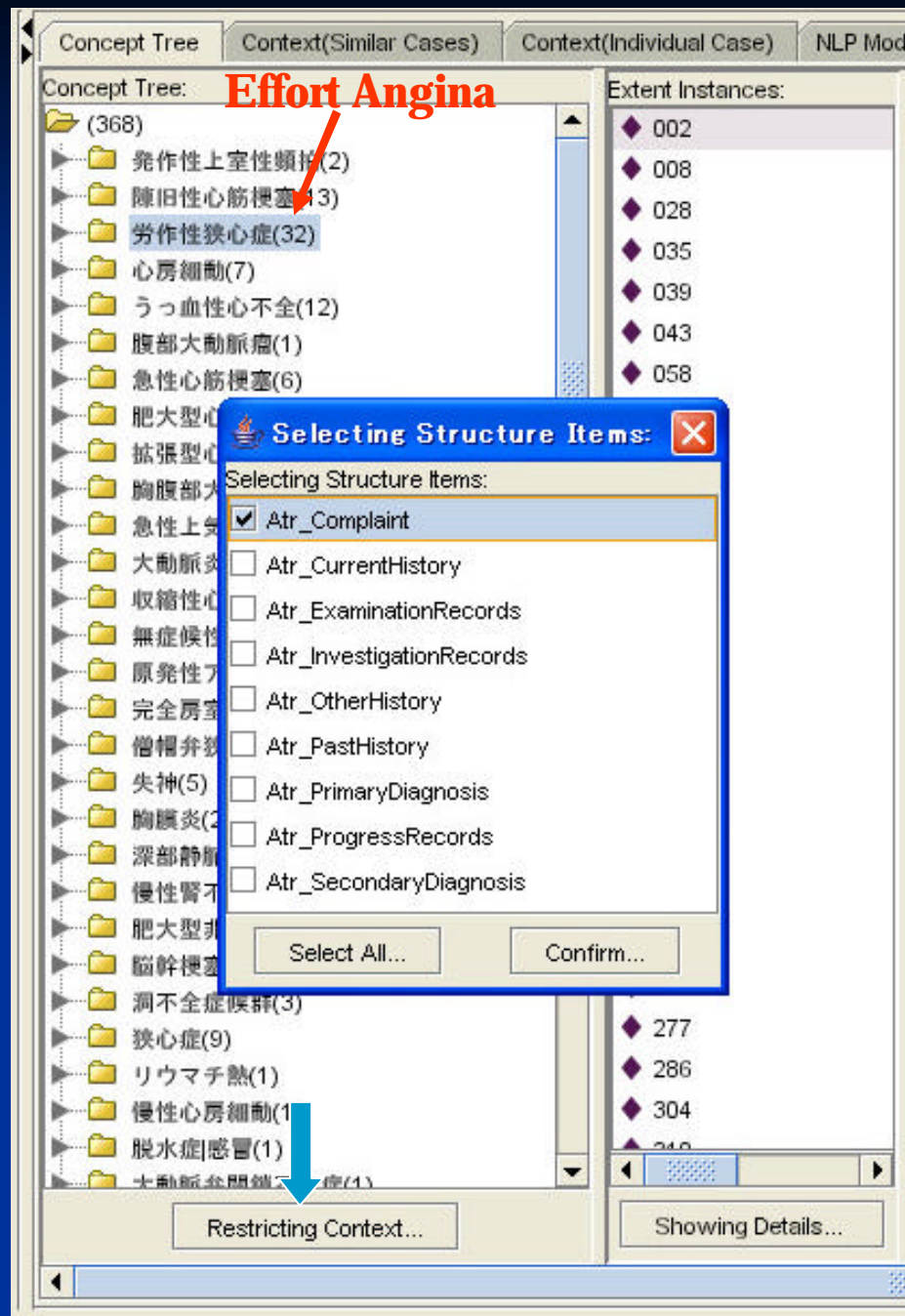
家族性高コレステロール血症(

depression.

肝機能障害.

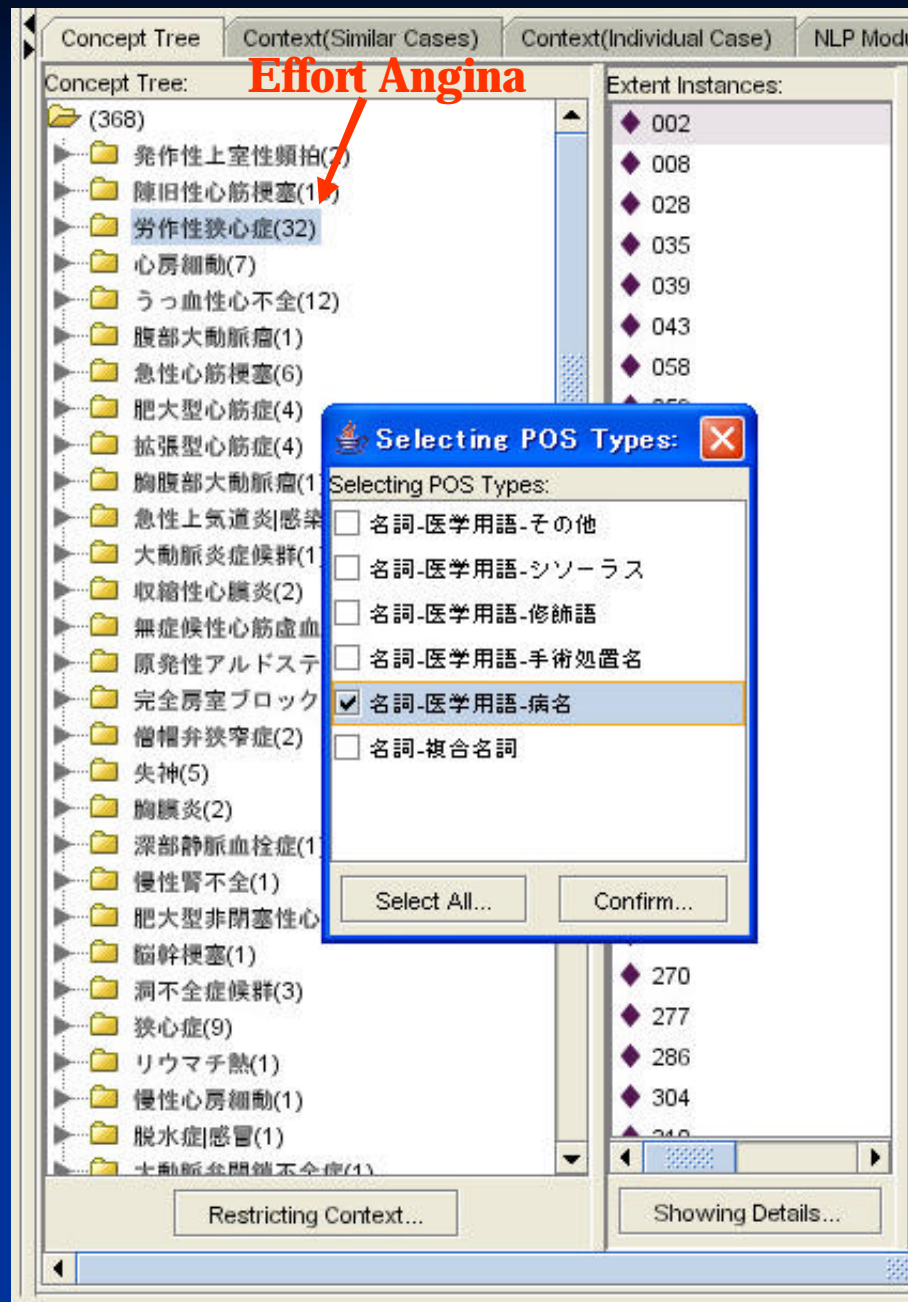
PastHistory

【既往歴】 25歳時:肺結核にて1年間入院 38歳時:肝機能障害 43歳時:扁桃摘出



✍ Context filter using Structure Knowledge

- ✍ Primary diagnosis
- ✍ Secondary diagnosis
- ✍ Chief complaints
- ✍ Current history
- ✍ Past history
- ✍ Other history
- ✍ Examination records
- ✍ Investigation records
- ✍ Progress records



✍ Context filter using
Semantic Knowledge of
medical concepts

✍ Noun-Diagnostic Term

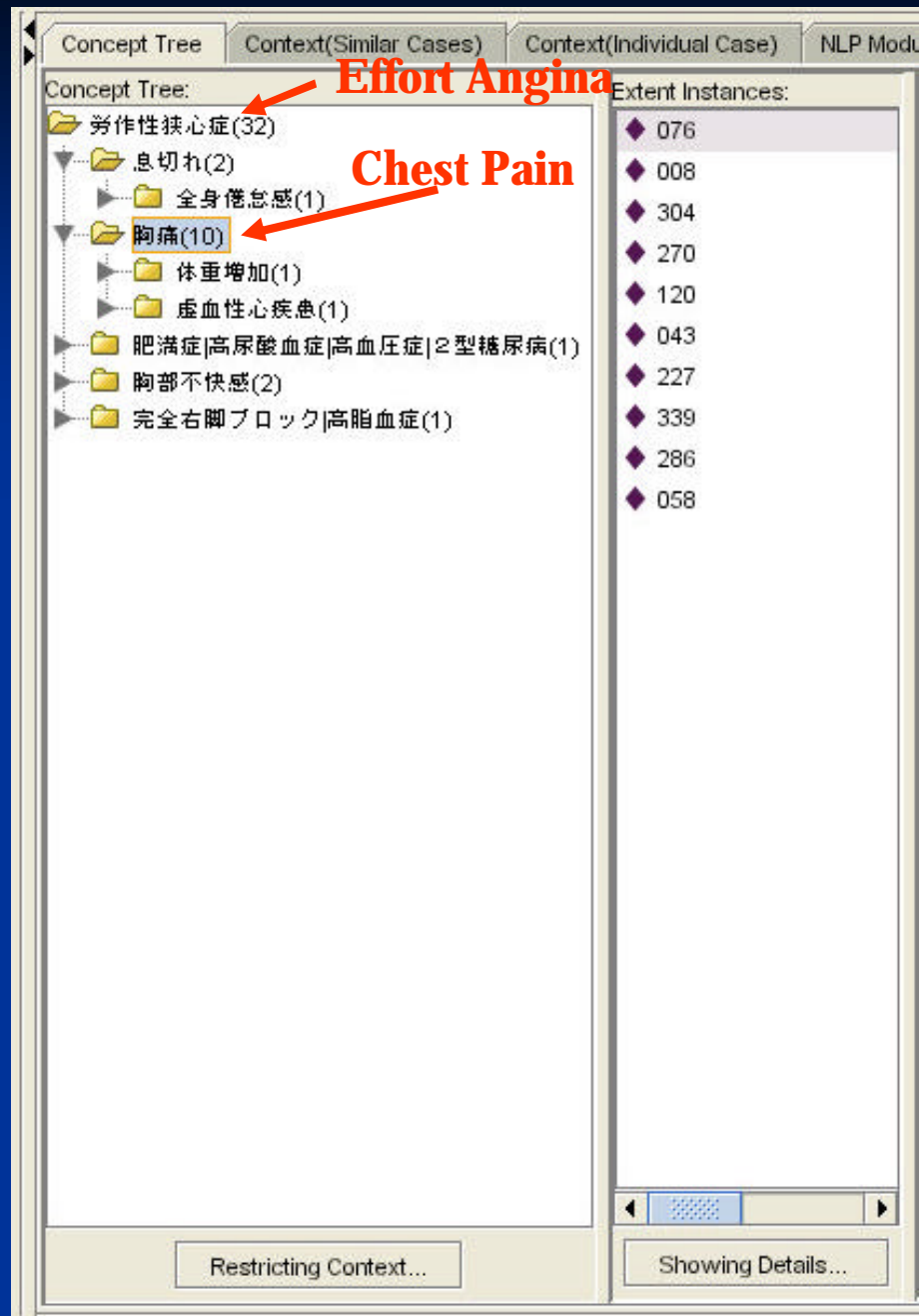
✍ Noun-Procedural Term

✍ Noun-Modifier

✍ Noun-MeSH Term

✍ Noun-Compound Term

✍ Noun-Others



✍ The filtered context:

✍ Formal objects:

✍ 32 cases with primary diagnosis of Effort Angina

✍ Formal attributes:

✍ All medical concepts with semantic type of Diagnosis Term, extracted from the Chief Complaints part of records

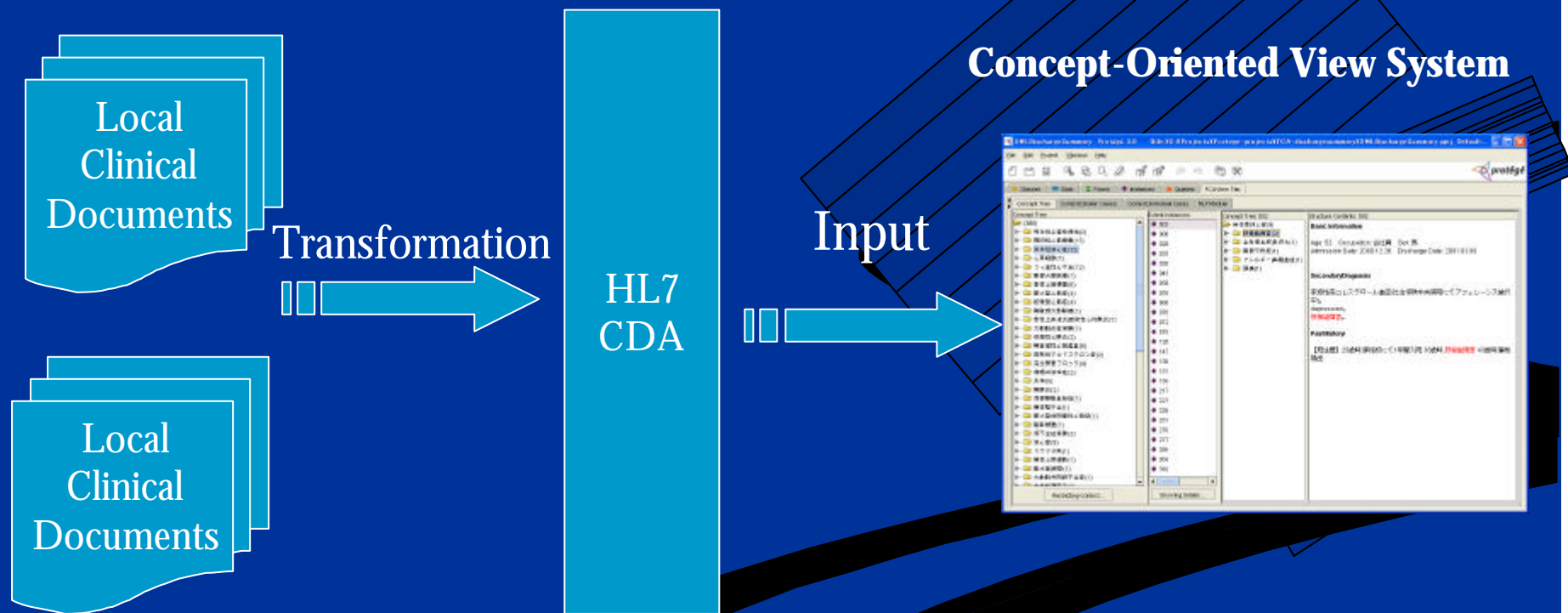
Discussion

- ✍ Automating view generation by integrating the NLP, Ontology-based knowledge base, with the FCA technique.
 - ✍ The individual patient records
 - ✍ The similar cases
- ✍ Features of the system
 - ✍ Processing free-text clinical data
 - ✍ Providing flexible way
 - ✍ for browsing using concept tree
 - ✍ for adapting the information needs of clinicians using context filters that are based on structural knowledge and semantic knowledge

Discussion

Generalization Consideration (I)

- ✍ To support HL-7 Clinical Data Architecture (CDA) as input format.



Discussion

Generalization Consideration (II)

- ✍ To visualize the Protégé-based ontologies.
- ✍ We assume that the instances (i.e. the formal objects) of a class in Protégé-2000, together with its own slots (i.e. the formal attributes), could form a formal context that would interests the users.
 - ✍ the instances of a class with its slots in Boolean type
 - ✍ the instances of a class with it slots in Multiple Instance type.



FCAView Tab plug-in provides [protégé-2000](#) users a novel way to visualize their ontologies using the **formal concept analysis (FCA)** technique.

FCA, a mathematical approach to data analysis based on lattice theory, has been advocated to support ontology building in several studies. Generally speaking, there are three elements to form a formal context: a set of formal objects, a set of formal attributes and the binary relations between them. The formal context is usually formed using a cross table, and graphically it also can be visualized using a lattice diagram. The two ways contain the same information about the formal context.

In the **FCAView Tab**, we hooked the protégé-2000 with an open source software [Concept Explorer](#) version 1.2. We assume that the instances (i.e. the formal objects) of a class in protégé-2000, together with its own slots (i.e. the formal attributes), could form a formal context that would interests the users. In particular, two kinds of formal context can be formed: 1) the instances of a class with its slots in Boolean type; 2) the instances of a class with it slots in Multiple Instance type.

Acknowledgements

- ✍ The study was supported in part by the Grants-in-Aid of the Japan Society for the Promotion of Science.
- ✍ Special thanks to Mr. Harold R. Solbrig in Division of BioMedical Informatics, Mayo Clinic, Rochester, MN, who contributed the idea about the generalization of our FCAView Tab.