XSEM – A Conceptual Model for XML

Martin Nečaský

Department of Software Engineering
Faculty of Mathematics and Physics, Charles University, Prague

2.11.2006

Outline

1) motivation

2) existing solutions

3) idea

4) XSEM model

5) future work

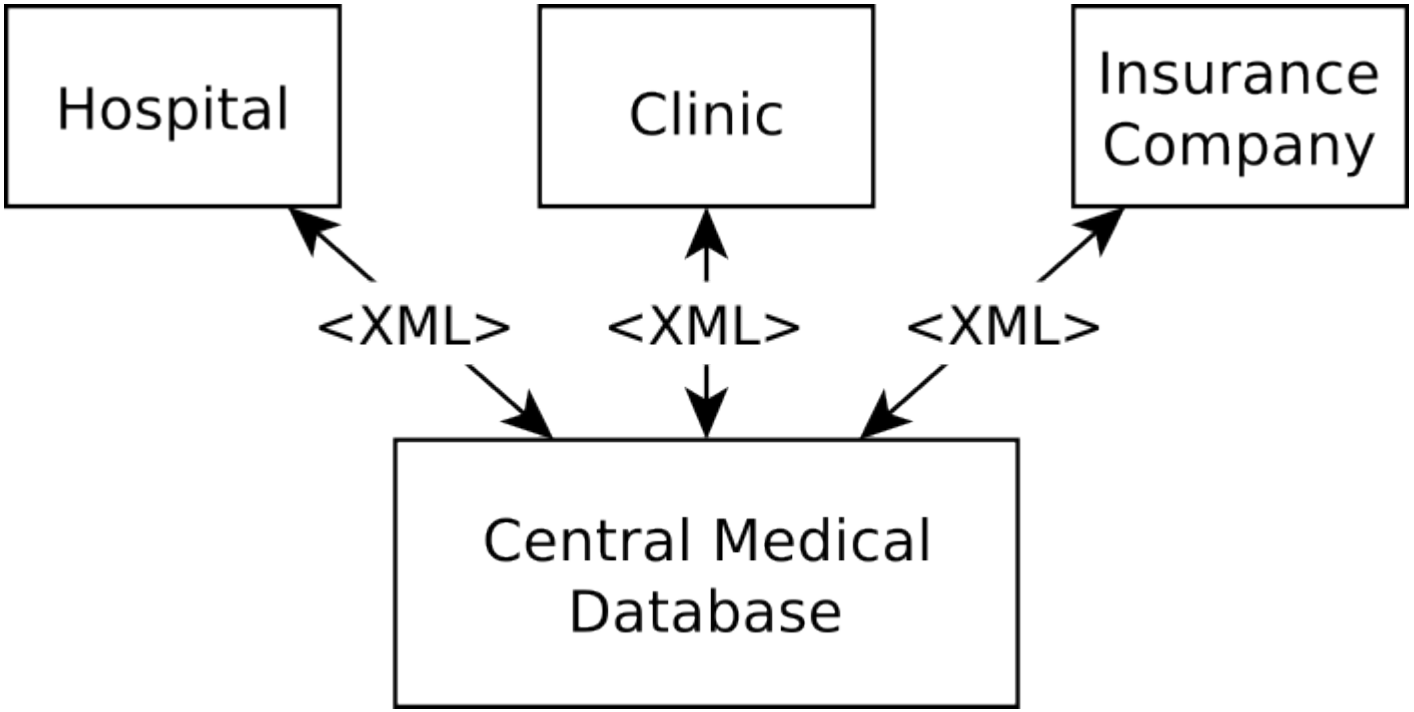
Motivation

- XML is applied as:
  - exchange data format
  - internal data representation format
  - logical database model
- XML appears on all application levels
- => modeling of XML data should become an inseparable part of the conceptual modeling of application and database data

Motivation - Example

- medical database collecting and managing data about patients from several external sources like hospitals, clinics, insurance companies, etc.
  - stores data in an internal representation
  - communicates with external sources through XML

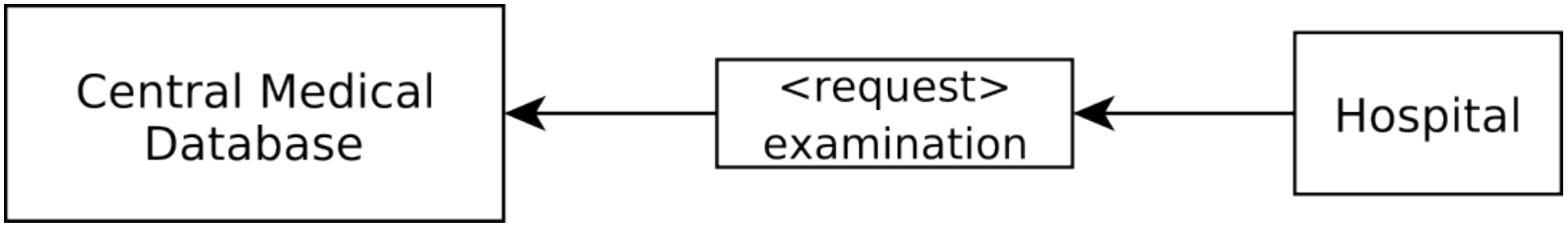
Motivation - Example



Motivation – Example Scenario

- physician in a hospital makes a diagnose of a patient
- to decide the diagnose, the physician needs results of some examinations of the patient
- physician requests the hospital system for the results
- hospital system requests the central medical database for the results

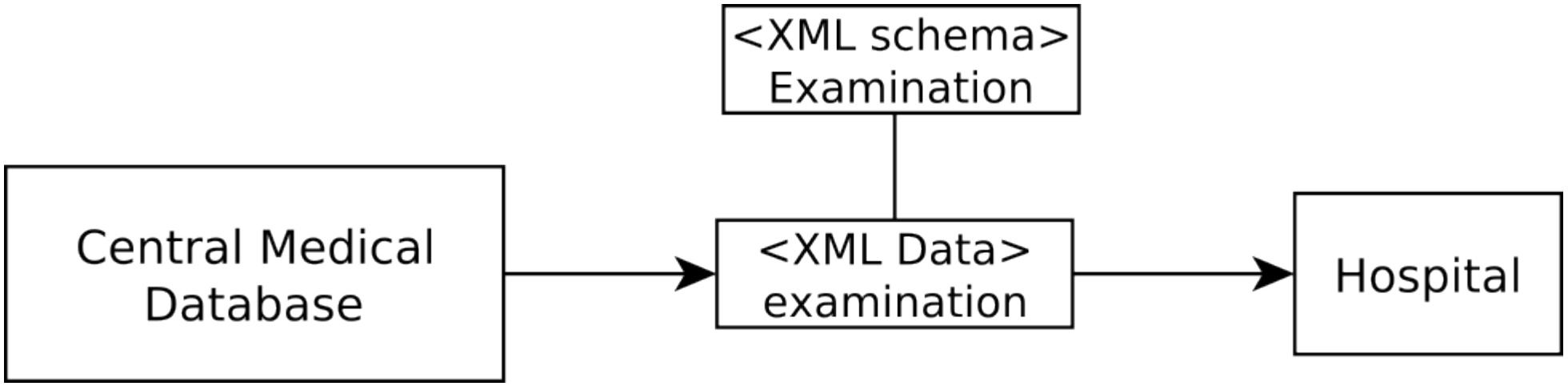
Motivation – Example Scenario



Motivation – Example Scenario

- central medical database exports the results from the internal representation into an XML document
- and sends the XML document with the results back to the hospital system
- hospital system receives the results and presents them to the physician

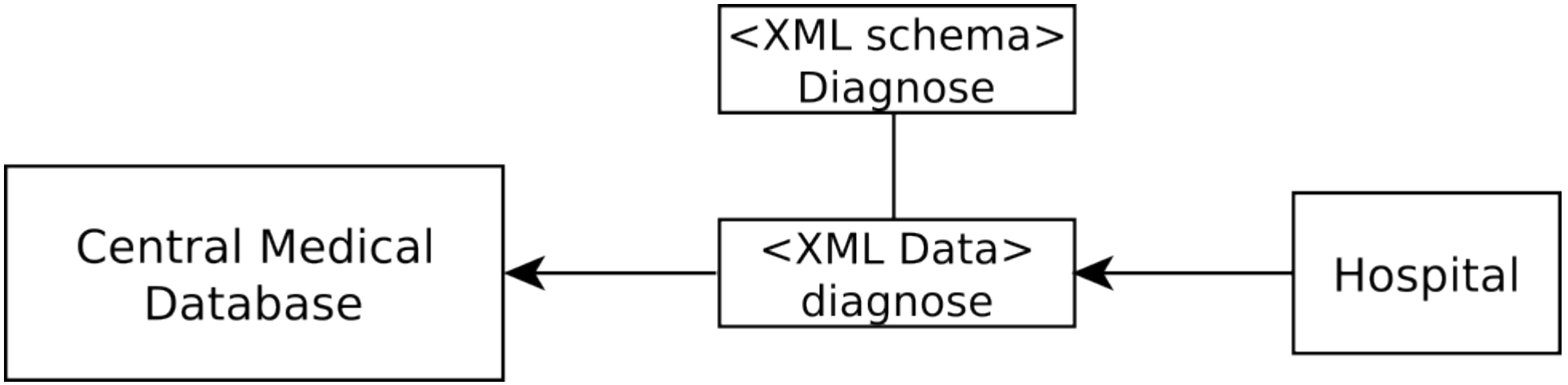
Motivation – Example Scenario



Motivation – Example Scenario

- on the base of the results, the physician diagnoses a disease of the patient and creates a report about the diagnose in the hospital system
- hospital system transforms the report into an XML document and sends it to the central medical database
- central medical database receives the XML document with the report and transforms it into its internal representation

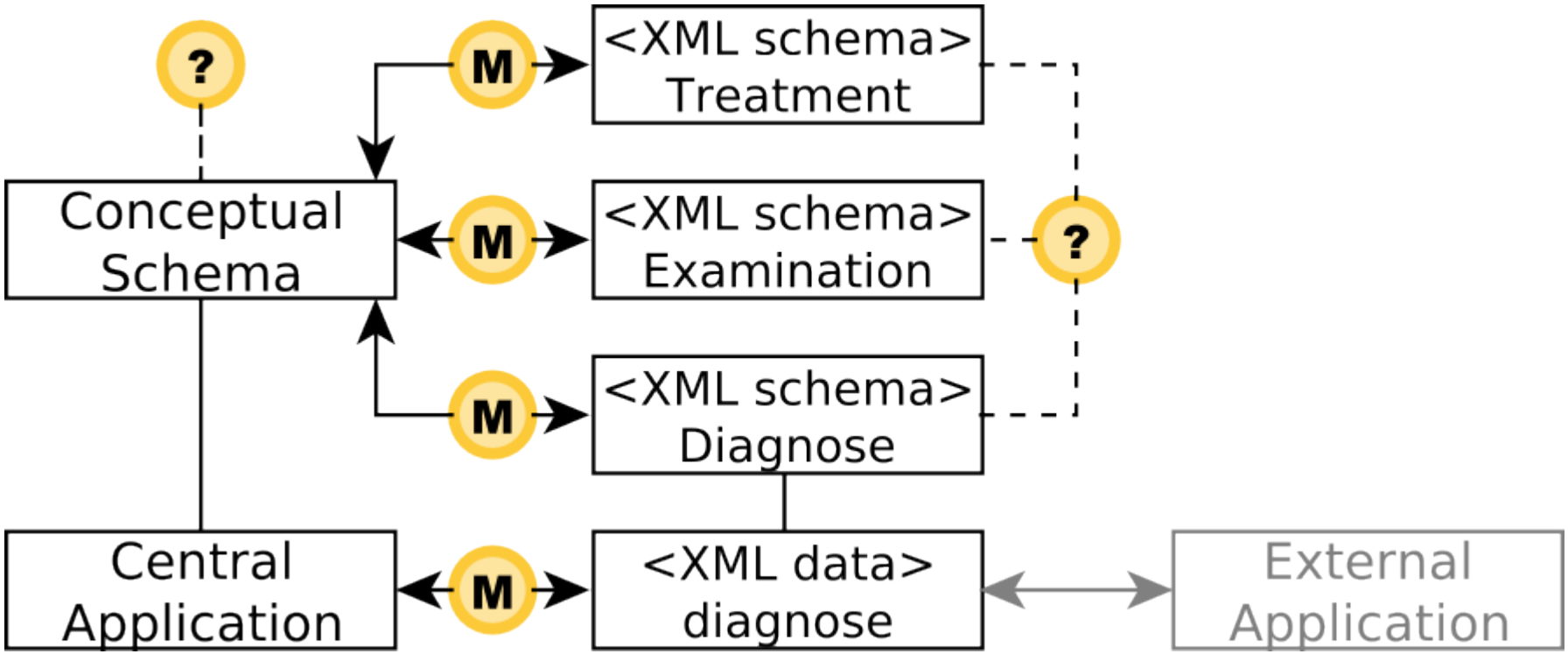
Motivation – Example Scenario



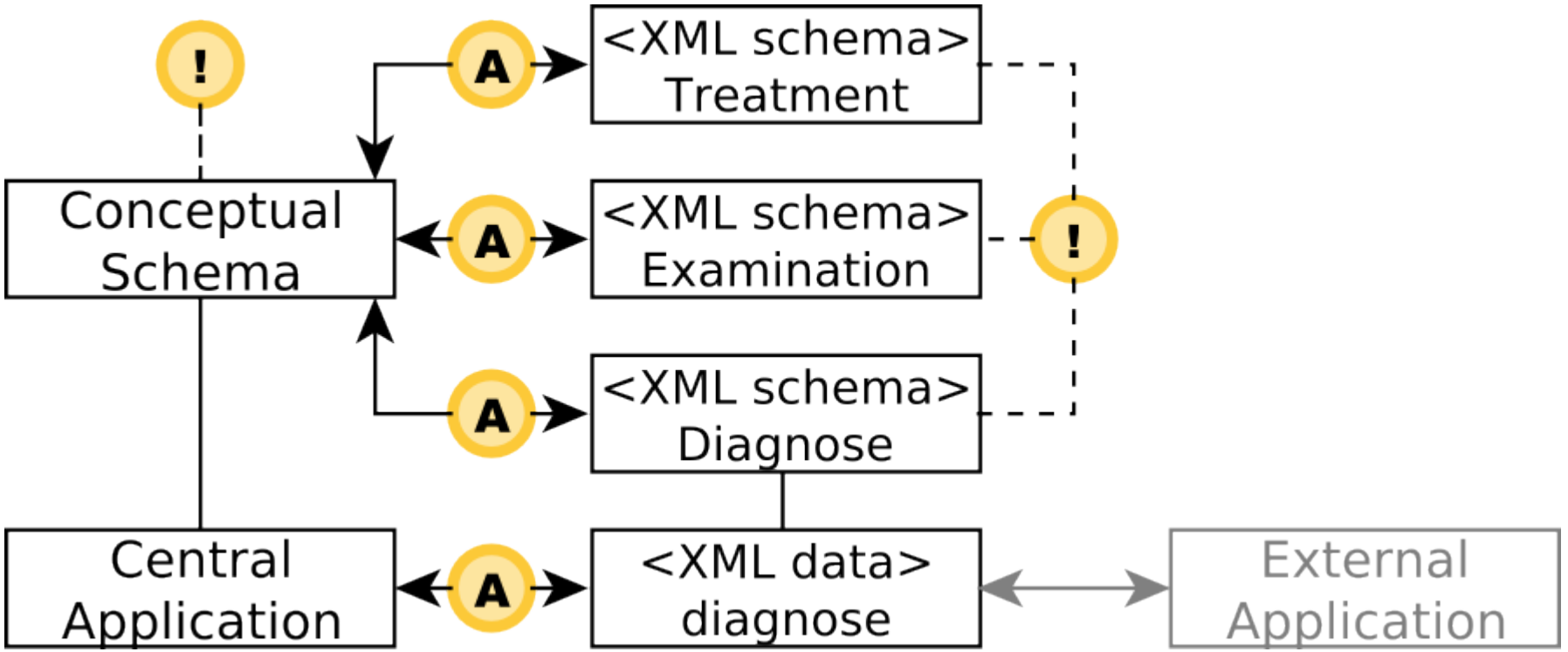
Motivation – Example

- connection of the example to the conceptual modeling:
  - we need to model the structure of XML documents used for the communication between the central medical database and external sources

Motivation – Current State



Motivation – Challenge



Outline

1) motivation
2) **existing solutions**
3) idea
4) XSEM model
5) future work

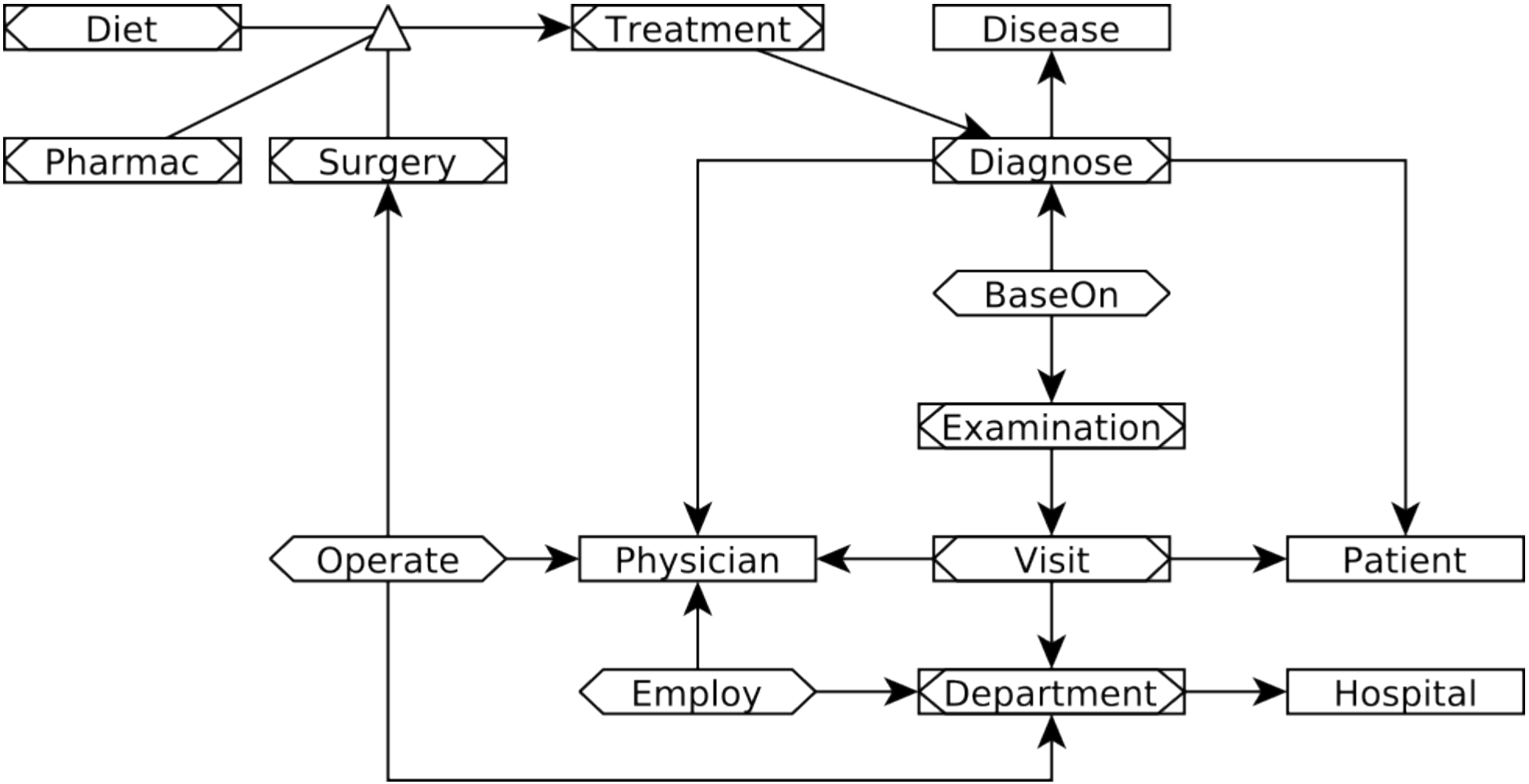
Special Features of XML

- when we model XML data we have to deal with some special features of XML:
  - hierarchical
  - ordered
  - irregular
  - semistructured (structured data mixed with an unstructured text)

Extending E-R

- E-R model is a popular, approved, and easy-to-use conceptual model
- idea: to extend the E-R model to be suitable for the modeling of the special XML features

Extending E-R

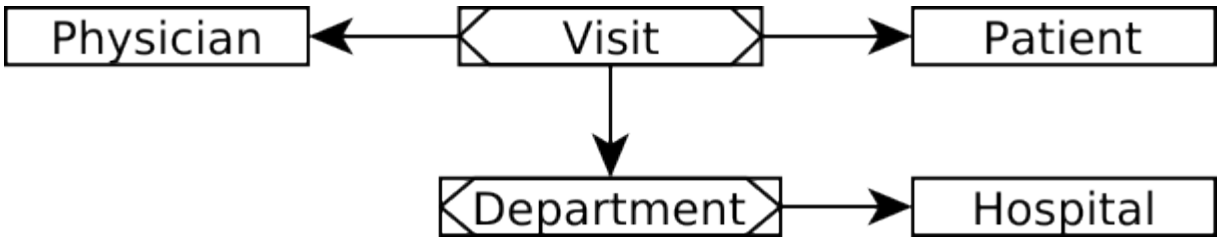


Extending E-R – Problems

- E-R is not hierarchical
  - *M:N* relationship types
  - *n*-ary relationship types
  - even if we consider a binary relationship type it is not clear how to organize it in a hiearchy
- existing solutions derive a hierarchical organization automatically without following user's requirements
- however, there can be many possible hierarchical organizations of the same data

Extending E-R – Problems

- possible hierarchical organization of *Visit*:
  - a list of visits
  - for each visit:
    - department
    - patient
    - physician



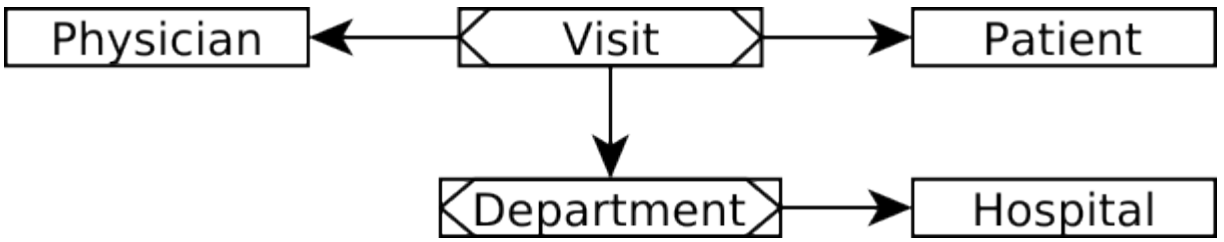
Extending E-R – Problems

- hierarchical organization of *Visit* in XML:

```
<visits>
   <visit><date>2006-09-12</date>
      <department><name>Department A</name>
         <hospital><name>Hospital B</name></hospital>
      </department>
      <patient><name>John Black</name></patient>
      <physician><name>Bill White</name></physician>
   </visit>
   ...
</visits>
```

Extending E-R – Problems

- another hierarchical organization of *Visit*:
  - a list of patiets
  - for each patient: a list of physicians the patient has visited
  - for each physician in the context of a patient: a list of visits of the physician made by the patient
  - for each visit in the context of a patient and a physician: a department where the visit was held



Extending E-R – Problems

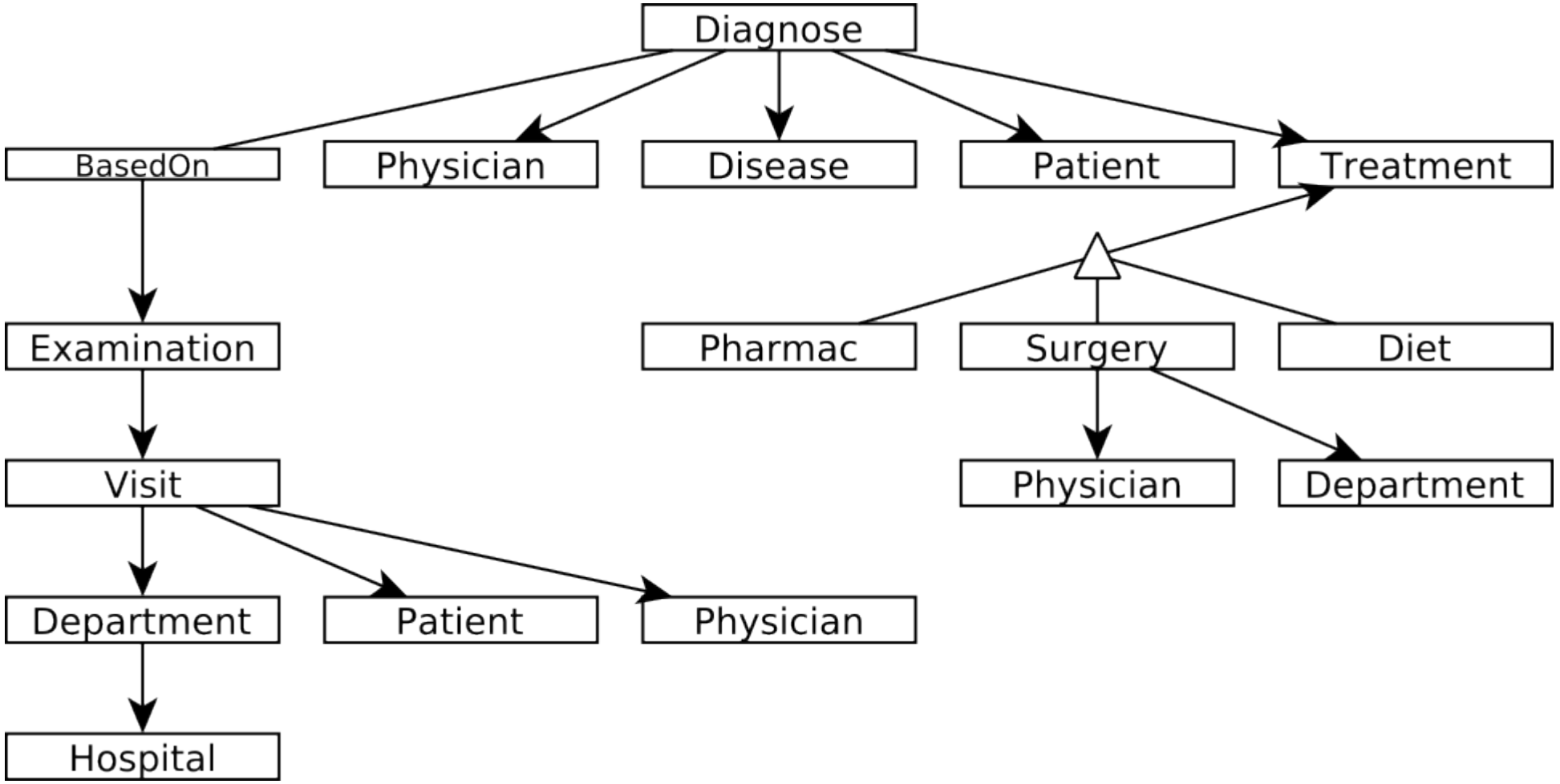
- hierarchical organization of V*isits* in XML

```
<patient><name>John Black</name>
  <physician><name>Bill White</name>
    <visit><date>2006-09-12</date>
      <department><name>Department A</name>
        <hospital><name>Hospital B</name></hospital>
      </department>
    </visit>
    ...
  </physician>
  ...
</patient>
```

Hierarchical Models

- idea: we can model data in hierarchies directly

Hierarchical Models

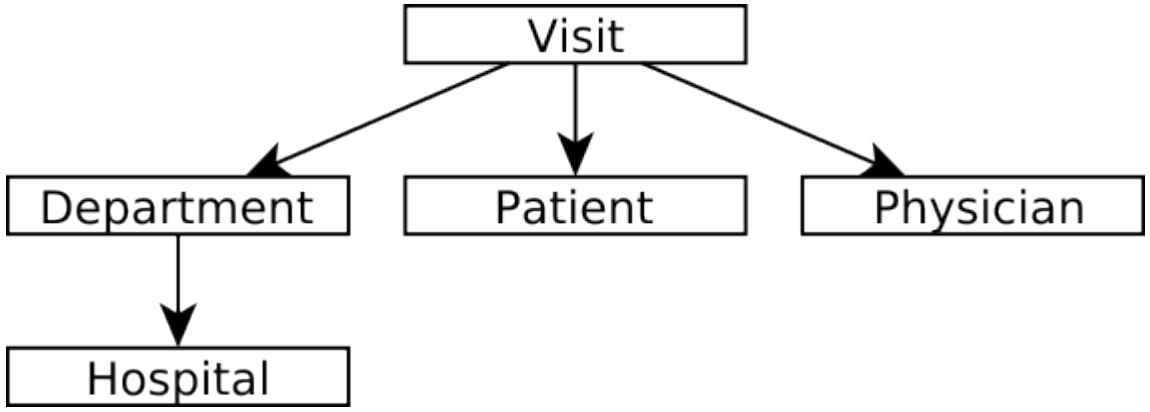


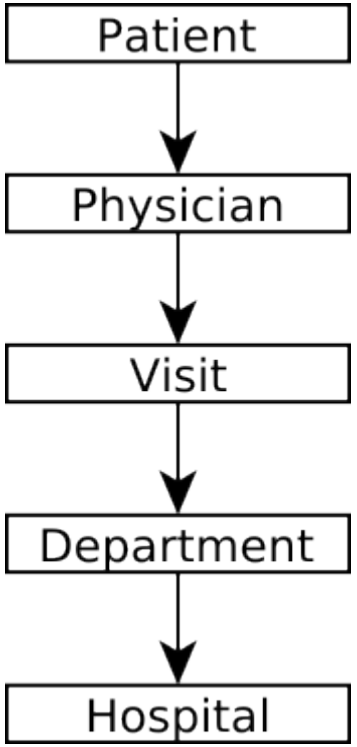
Hierarchical Models – Problems

- not so clear and much more complex schemes
- captured domain semantics is poor
- hierarchical schema models only one of the possible hierarchical organization of the same data
  - there can be more than this one
  - if we want to model more hierarchical organizations of the same data we have to interconnect parts of schemes representing the same data in some way

Hierarchical Models – Problems

- two different schemes modeling the same data in different hierarchical organizations:





Outline

1) motivation
2) existing solutions
3) **idea**
4) XSEM model
5) future work

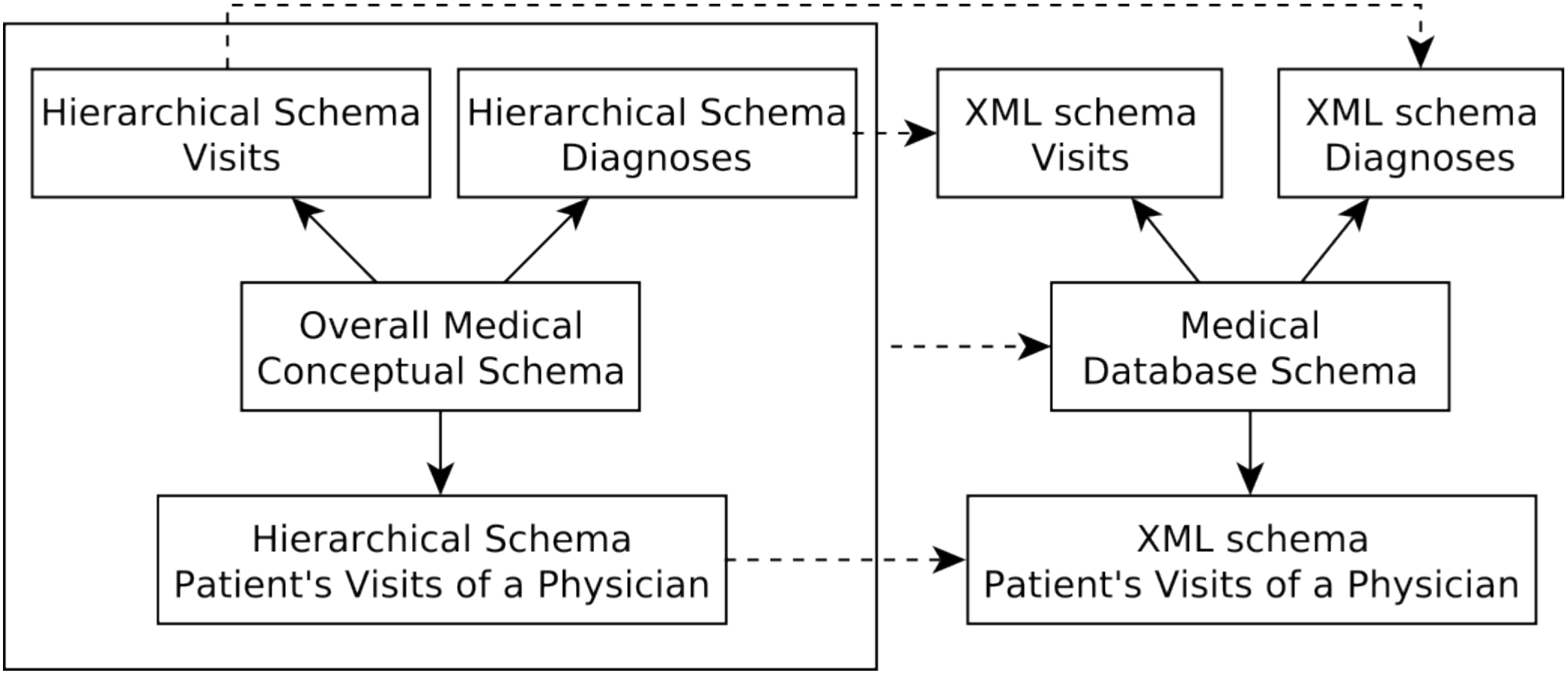
Idea – Logical Level

- internal logical schema
- users access data in the internal representation through XML documents described by the XML schemes
- XML schemes as hierarchical views on parts of the internal logical schema
- each group of users needs different structure of XML documents containing the same data
  - different hierarchical views (XML schemes) on the same parts of the internal logical schema

Idea – Conceptual Level

- overall conceptual schema describing data
- hierarchical conceptual views derived from the overall conceptual schema describing required XML schemes

Idea


Hierarchical Schema
Visits
Hierarchical Schema
Diagnoses
XML schema
Visits
XML schema
Diagnoses
Overall Medical
Conceptual Schema
Medical
Database Schema
Hierarchical Schema
Patient's Visits of a Physician
XML schema
Patient's Visits of a Physician

Outline

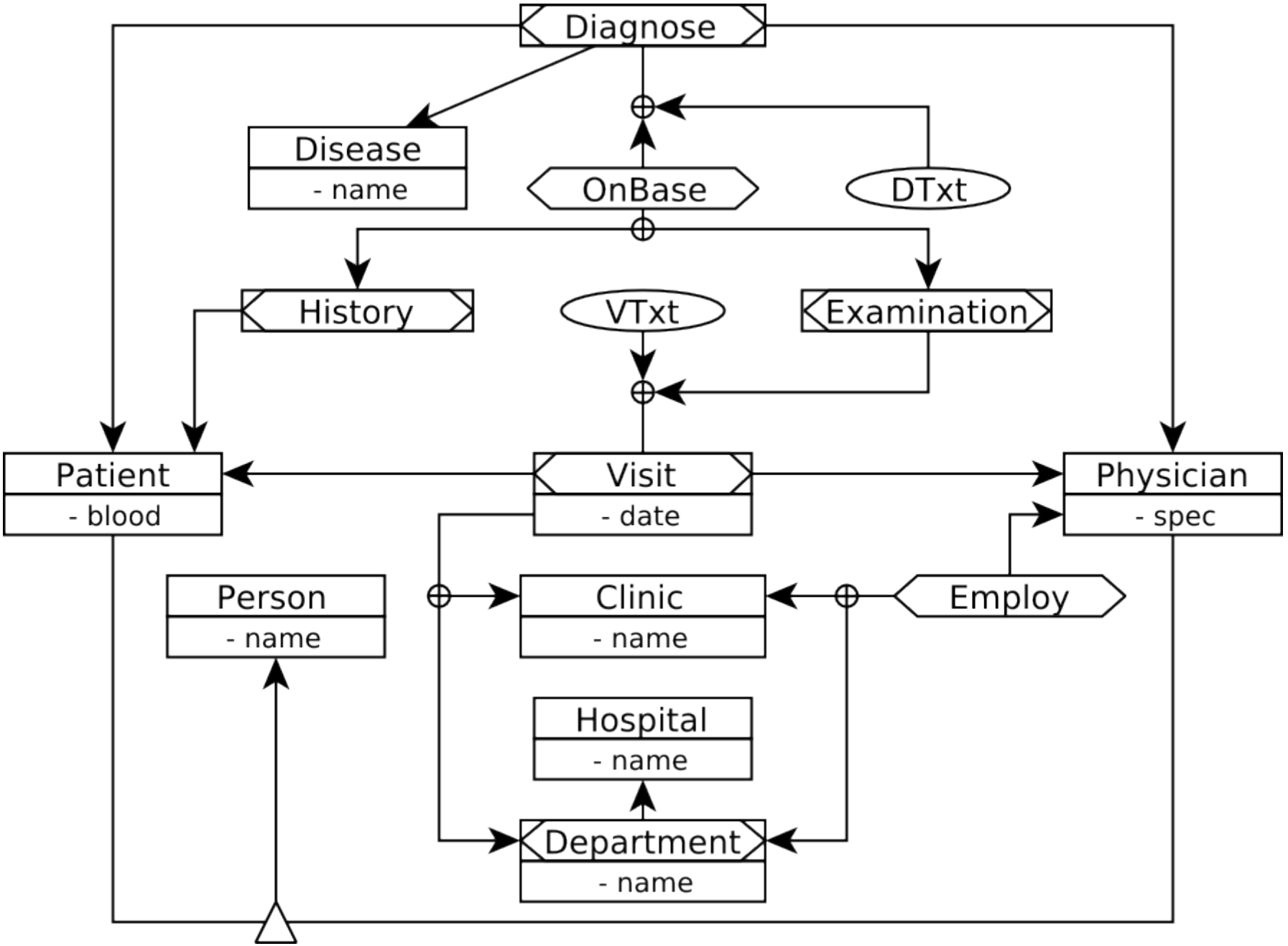
1) motivation
2) existing solutions
3) idea
4) **XSEM model**
5) future work

XSEM

- conceptual model for XML based on the previous idea
- devides a conceptual modeling process to two parts:
  - the first part consits of designing an overall conceptual schema of a domain using *XSEM-ER*
  - the second part consists of designing hierarchical organizations of the structures from the first part using *XSEM-H*

XSEM-ER

- extension of E-R proposed by Chen
- modeling of the special XML features:
  - irregular structure (*cluster types*)
  - ordering (*ordering constraints*)
  - semistructured data (*data node types* and *cluster types*)
- hierarchical organization is not important here


Diagnose
Disease
- name
OnBase
DTxt
History
VTxt
Examination
Patient
- blood
Visit
- date
Physician
- spec
Person
- name
Clinic
- name
Employ
Hospital
- name
Department
- name

XSEM-ER

- irregular structure in XML

```
<patient><name>John Black</name>
  <physician><name>Bill White</name>
    <visit><date>2006-09-12</date>
      <department><name>Department A</name>
        <hospital><name>Hospital B</name></hospital>
      </department>
    </visit>
    <visit><date>2006-09-26</date>
      <clinic><name>Clinic C</name></clinic>
    </visit>
  </physician> ... </patient>
```

XSEM-ER

- semistructured data in XML

```
<visit><date>2006-09-12</date>
   <clinic><name>Clinic C</name></clinic>
   <patient><name>John Black</name></patient>
   <physician><name>Bill White</name></physician>
   <course>Because of the family anamnesis
     (<anamnesis>...</anamnesis>) there is a suspicion of
     liver cancer. Consequently, I made a laboratory
     examinations (<examination>...</examination> and
     <examination>...</examination>) ... </course>
</visit>
```

XSEM-ER – Hierarchical Projections

- step between the non-hierarchical XSEM-ER level and the hierarchical XSEM-H level
- binarization of relationship types and weak entity types
- for example, *Visit* can be represented by the following hierarchy:
  - a list of patients
  - for each patient: a list of patient's visits
  - for each patient's visit: the visited physician and the department or clinic where the patient visited the physician

XSEM-ER – Hierarchical Projections

- the hierarchy describes the structure of the following example XML document:

```
<patient><name>John Black</name>
   <visit><date>2006-09-12</date>
      <physician><name>Bill White</name></physician>
      <department><name>Department A</name>
         <hospital><name>Hospital B</name></hospital>
      </department>
   </visit>
   <visit><date>2006-10-11</date>
      <physician><name>Jack Brown</name></physician>
      <clinic><name>Clinic D</name></clinic>
   </visit> ...
</patient>
```

XSEM-ER – Hierarchical Projections

- the hierarchy is formally described by the following hierarchical projections:

$$Visit[Patient \rightarrow Visit]$$

$$Visit^{Patient}[Visit \rightarrow Physician]$$

$$Visit^{Patient}[Visit \rightarrow Department+Clinic]$$

XSEM-ER – Hierarchical Projections

- the hierarchical projections:

Visit[Visit → Patient]

Visit[Visit → Physician]

Visit[Visit → Department+Clinic]

XSEM-ER – Hierarchical Projections

- describe the structure of the following:

```
<visits>
   <visit><date>2006-09-12</date>
      <department><name>Department A</name>
         <hospital><name>Hospital B</name></hospital>
      </department>
      <patient><name>John Black</name></patient>
      <physician><name>Bill White</name></physician>
   </visit>
   <visit><date>2006-10-11</date>
      <clinic><name>Clinic D</name></clinic>
      <patient><name>John Black</name></patient>
      <physician><name>Jack Brown</name></physician>
   </visit> ... </visits>
```

XSEM-ER – Hierarchical Projections

- and the hierarchical projections:

$$Visit[Patient \rightarrow Physician]$$

$$Visit^{Patient}[Physician \rightarrow Visit]$$

$$Visit^{Patient\ Physician}[Visit \rightarrow Department{+}Clinic]$$

XSEM-ER – Hierarchical Projections

- describe the structure of the following

```
<patient><name>John Black</name>
   <physician><name>Bill White</name>
      <visit><date>2006-09-12</date>
         <department><name>Department A</name>
            <hospital><name>Hospital B</name></hospital>
         </department>
      </visit> ... </physician>
   <physician><name>Jack Brown</name>
      <visit><date>2006-10-11</date>
         <clinic><name>Clinic D</name></clinic>
      </visit> ... </physician>
   ...
</patient>
```

XSEM-ER – Hierarchical Projections

- we can specify integrity constraints for hierarchical projections:
  - ordering constraints
  - cardinality constraints

XSEM-H

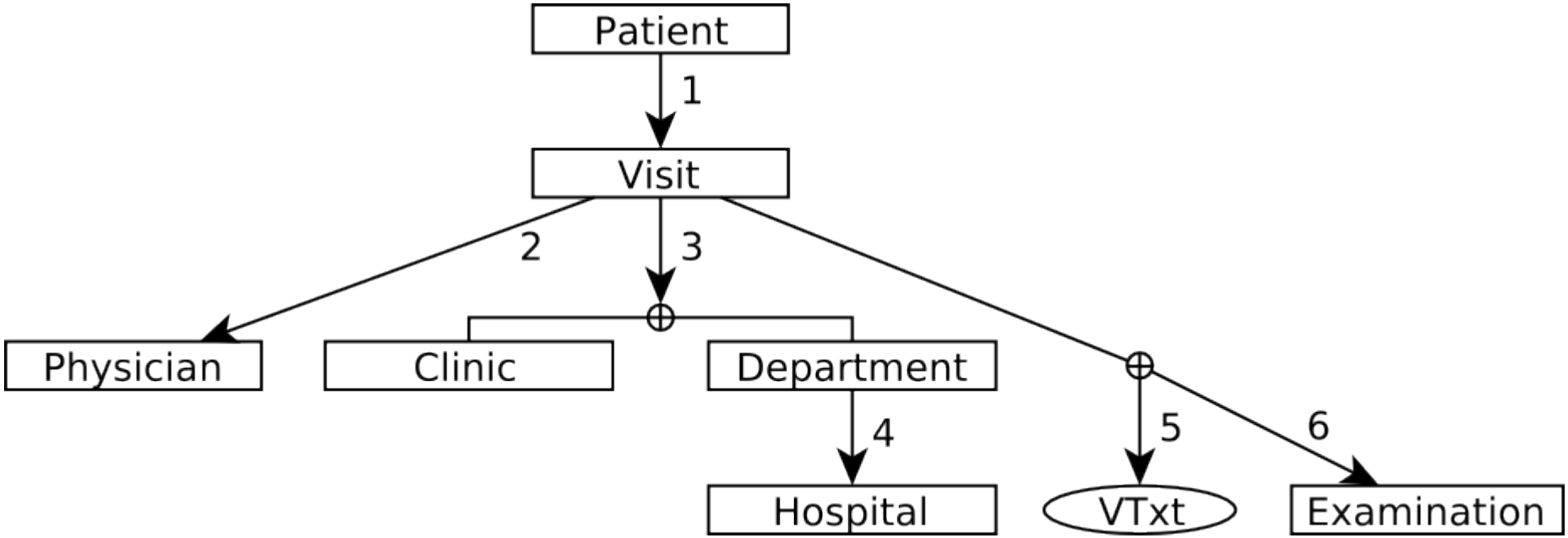
- specification of a hierarchical organization of a part of a given XSEM-ER schema using hierarchical projections
- does not add any semantics

XSEM-H – Schema

- derived from an XSEM-ER schema by ***transformation operators***
- oriented graph
- nodes represent entity types, relationship types, and data node types from the XSEM-ER schema
- edges represent hierarchical projections of weak entity types and relationship types from the XSEM-ER schema and references

XSEM – Tranformation Operators

- user does not specify hierarchical projections to compose an XSEM-H schema
- he or she uses transformation operators generating hierarchical projections according to supplied parameters



1: $Visit[Patient \rightarrow Visit]$

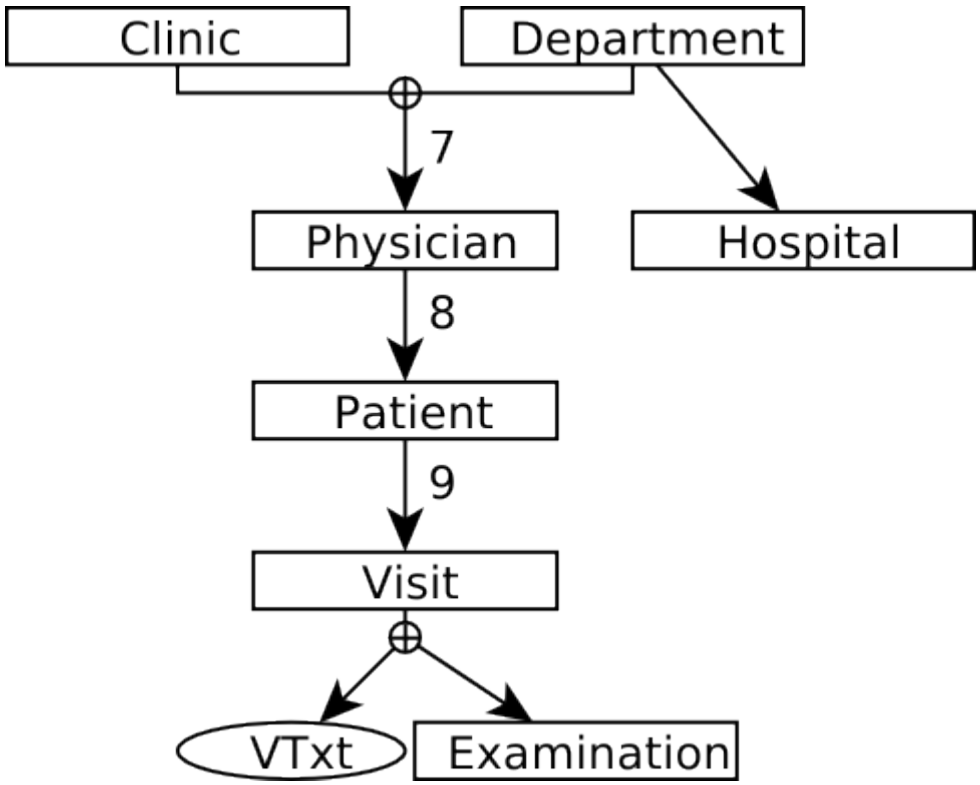
2: $Visit^{Patient}[Visit \rightarrow Physician]$

3: $Visit^{Patient}[Visit \rightarrow Department+Clinic]$

4: $Department[Department \rightarrow Hospital]$

5: $VTxt[Visit \rightarrow VTxt]$

6: $Examination[Visit \rightarrow Examination]$

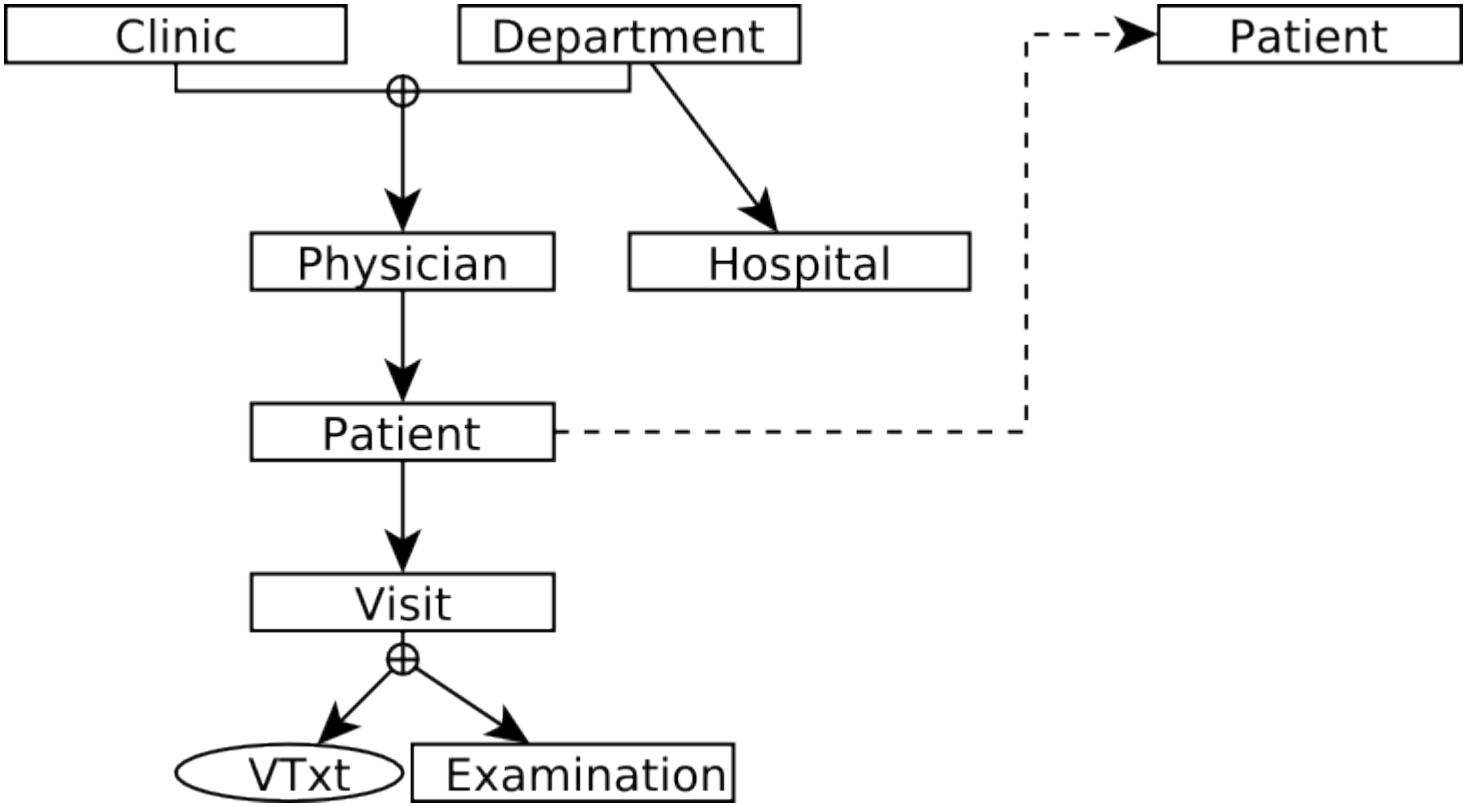


7: *Visit[Department+Clinic→ Physician]*

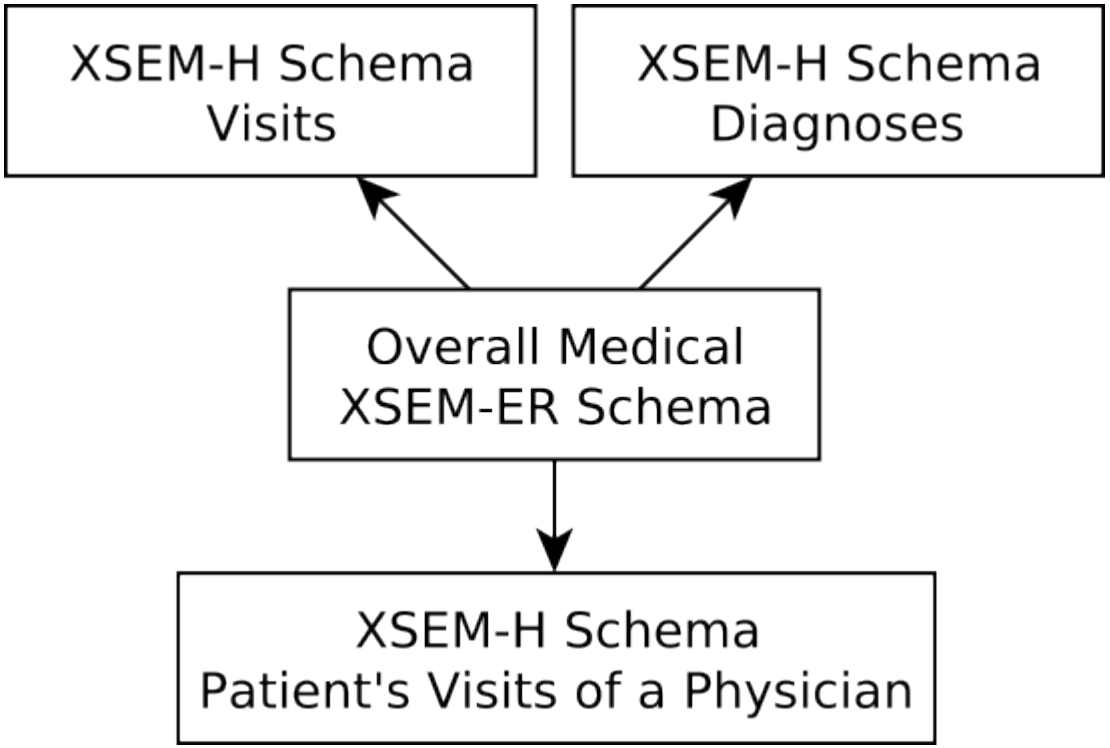
8: *Visit*$^{Department+Clinic}$*[Physican → Patient]*

9: *Visit*$^{Department+Clinic\ Physician}$*[Patient → Visit]*

XSEM – References



XSEM



- hierarchical projections are the connection between an XSEM-ER schema and derived XSEM-H schemes

Outline

1) motivation
2) existing solutions
3) idea
4) XSEM model
5) future work

Future Work

- translation to the XML schema level
  - grammar-based languages (XML Schema, Relax NG) for describing structure
  - pattern-based languages (Schematron, XSLT, XQuery) for describing more complex integrity constraints
- translation to the logical database level
  - (object-)relational schemes
  - xml schemes
  - hybrid schemes

Publications

- Necasky M.: Conceptual Modeling for XML: A Survey. Technical Report No. 2006-3, Dep. of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague, 2006, 54 p.
- Necasky M.: XSEM - A Conceptual Model for XML Data. Databases and Information Systems. Seventh International Baltic Conference on Databases and Information Systems. Communications, Materials of Doctoral Consortium. O. Vasilecas, J. Eder, A. Caplinskas (Eds.). Vilnius: Technika, July 2006, s. 328-331. ISBN 9955-28-013-1.
- **Necasky M.: XSEM - A Conceptual Model for XML Data. Accepted for APCCM 2007, Ballarat, Victoria, Australia, January 2007, 12 p.**