# Improving the Usability of HL7 Information Models by Automatic Filtering

Antonio Villegas[1] Antoni Olivé[1] Josep Vilalta[2]

[1]Services and Information Systems Engineering Department
Universitat Politècnica de Catalunya

[2]HL7 Education & e-Learning Services
HL7 Spain (Health Level Seven International)

ESSI Seminar
June 2, 2010

# Outline

1. Introduction
   - IEEE 6th World Congress on Services (SERVICES 2010)
2. Health Level Seven
   - Healthcare Services
   - Reference Models Overview
   - RIM
   - D-MIM
   - R-MIM
3. Filtering HL7 Models
   - Overview
   - User Preferences
   - Filtering Measures
   - Interest Set
   - Filtered Information Model
4. Evaluation
   - Precision Analysis
   - Time Analysis
5. Conclusions

# Outline

1. Introduction
   - IEEE 6th World Congress on Services (SERVICES 2010)
2. Health Level Seven
   - Healthcare Services
   - Reference Models Overview
   - RIM
   - D-MIM
   - R-MIM
3. Filtering HL7 Models
   - Overview
   - User Preferences
   - Filtering Measures
   - Interest Set
   - Filtered Information Model
4. Evaluation
   - Precision Analysis
   - Time Analysis
5. Conclusions

# Introduction

## Improving the Usability of HL7 Information Models by Automatic Filtering

Antonio Villegas and Antoni Olivé
*Services and Information Systems Engineering Department*
*Universitat Politècnica de Catalunya*
*Barcelona, Spain*
*Email: {avillegas, olive}@essi.upc.edu*

Josep Vilalta
*HL7 Education & e-Learning Services*
*HL7 Spain (Health Level Seven International)*
*Barcelona, Spain*
*Email: jvilalta@vico.org*

***Abstract*—The amount of knowledge represented in the Health Level 7 International (HL7) information models is very large. The sheer size of those models makes them very useful for the communities for which they are developed. However, the size of the models and their overall organization makes it difficult to manually extract knowledge from them. We propose to extract that knowledge by using a novel filtering method that we have developed. Our method is based on the concept of class interest as a combination of class importance and class closeness. The application of our method automatically obtains a filtered information model of the whole HL7 models according to the user preferences. We show that the use of a prototype tool that implements that method and produces such filtered model improves the usability of the HL7 models due to its high precision and low computational time.**

***Keywords*-Usability, Health Level Seven International, HL7, Models, Filtering, UML**

### I. Introduction

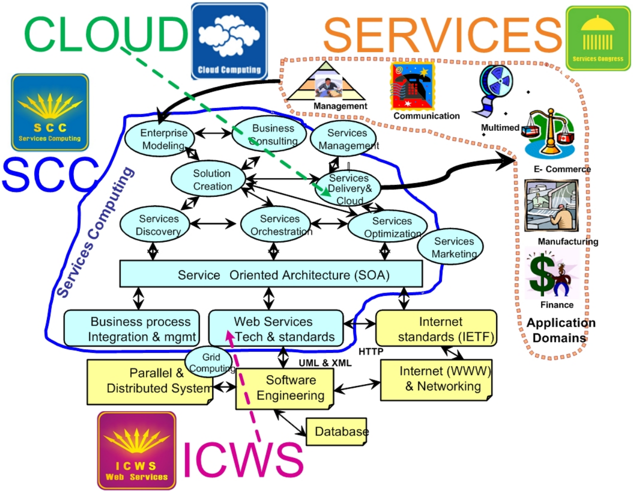
The Health Level Seven International (HL7) is a not-for- … to manually extract knowledge from them. This problem is shared by other large models [6].

Currently, there is a lack of computer support to make those models usable for the goal of knowledge extraction. In this paper, we propose to extract that knowledge by using a novel filtering method that we have developed, and we show that the use of our prototype implementation of that method improves the usability of HL7 information models.

The structure of the paper is as follows. Section II introduces the HL7 models and describes the main UML constructs used to build them. Section III describes the concept of class importance and references the methods that can be used to compute it. Section IV describes the concept of class interest with respect to a focus set of classes and explains how to compute it. Section V presents our model filtering method. Section VI evaluates the use of the method in the context of the HL7 models. Finally, Section VII summarizes the conclusions and points out future work.

# IEEE 6th World Congress on Services (SERVICES 2010)

www.servicescongress.org/2010 July 5–10 Miami USA



# IEEE 6th World Congress on Services (SERVICES 2010)

www.servicescongress.org/2010 July 5–10 Miami USA

Business services sectors:

- Advertising Services
- Banking Services
- Broadcasting & Cable TV Services
- Business Services
- Casinos & Gaming Services
- Communications Services
- Cross-industry Services
- Design Automation Services
- Energy and Utilities Services
- Financial Services
- Government Services
- Healthcare Services
- Hotels & Motels Services
- Insurance Services
- Internet Services
- Motion Pictures Services
- Personal Services
- Printing & Publishing Services
- Real Estate Operations Services
- Recreational Activities Services
- Rental & Leasing Services
- Restaurants Services
- Retail Services
- Schools and Education Services
- Security Systems & Services
- Technology Services
- Travel and Transportation Services
- Waste Management Services
- Wholesale Distribution Services

# IEEE 6th World Congress on Services (SERVICES 2010)

www.servicescongress.org/2010 July 5–10 Miami USA

Business services sectors:

- Advertising Services
- Banking Services
- Broadcasting & Cable TV Services
- Business Services
- Casinos & Gaming Services
- Communications Services
- Cross-industry Services
- Design Automation Services
- Energy and Utilities Services
- Financial Services
- Government Services
- Healthcare Services
- Hotels & Motels Services
- Insurance Services
- Internet Services
- Motion Pictures Services
- Personal Services
- Printing & Publishing Services
- Real Estate Operations Services
- Recreational Activities Services
- Rental & Leasing Services
- Restaurants Services
- Retail Services
- Schools and Education Services
- Security Systems & Services
- Technology Services
- Travel and Transportation Services
- Waste Management Services
- Wholesale Distribution Services

# Outline

1. Introduction
   - IEEE 6th World Congress on Services (SERVICES 2010)
2. Health Level Seven
   - Healthcare Services
   - Reference Models Overview
   - RIM
   - D-MIM
   - R-MIM
3. Filtering HL7 Models
   - Overview
   - User Preferences
   - Filtering Measures
   - Interest Set
   - Filtered Information Model
4. Evaluation
   - Precision Analysis
   - Time Analysis
5. Conclusions

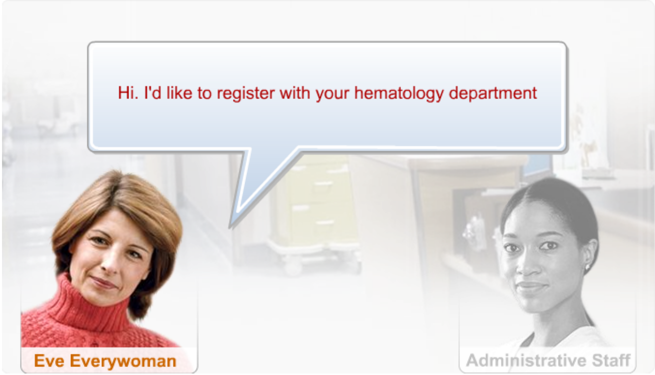
# Healthcare Services

## Example*


* pictures from `hl7.org`

# Healthcare Services

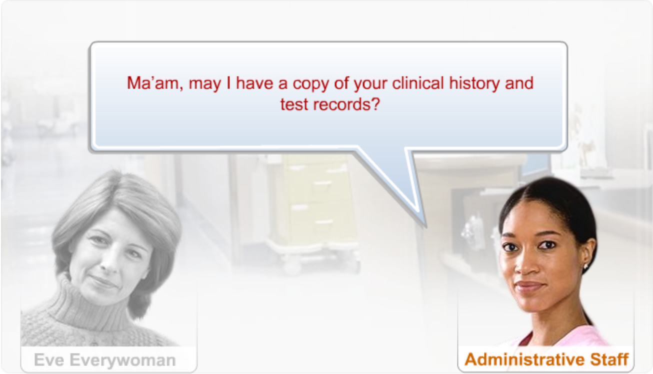
## Example*




* pictures from hl7.org

# Healthcare Services

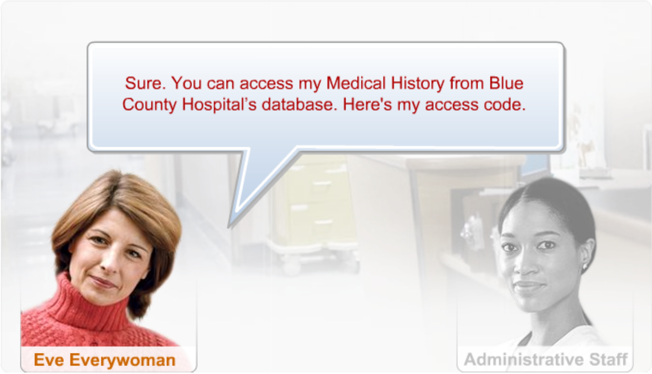
## Example*




* pictures from hl7.org

# Healthcare Services

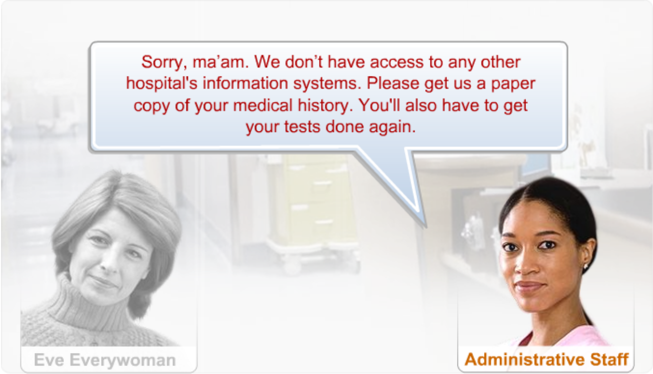
## Example*




* pictures from `hl7.org`

# Healthcare Services

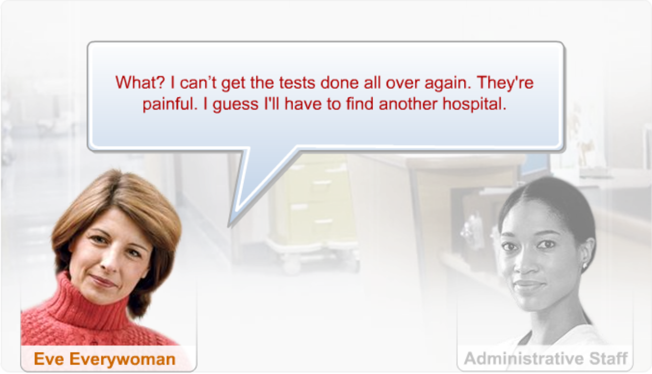
## Example*




* pictures from `hl7.org`

# Healthcare Services

## Example*




* pictures from `hl7.org`

# Healthcare Services

Key challenges faced by healthcare organizations today include:

- Impact on the safety, effectiveness, and cost of healthcare by not having the **right information** at the **right place** at the **right time**.
- Presentation of **disparate healthcare information** at the point of treatment
- Increased cost in **transferring paper** records in this age of e-commerce

# Healthcare Services

## Motivation

### Problem

Inability to share and manage data within and across organizations

### Requirement

Interoperability of Services

### Solution

Use of Standards

# Healthcare Services
## Motivation

**Problem**

Inability to share and manage data within and across organizations

**Requirement**

Interoperability of Services

**Solution**

Use of Standards

# Healthcare Services

## Motivation

**Problem**

Inability to share and manage data within and across organizations

**Requirement**

Interoperability of Services

**Solution**

Use of Standards

# Healthcare Services
## Motivation

### Problem
Inability to share and manage data within and across organizations

### Requirement
Interoperability of Services

### Solution
Use of Standards

# Health Level Seven

The Health Level Seven International (HL7) is a not-for profit, ANSI-accredited **standards developing organization** dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of **health services**.

HL7 develops specifications, the most widely used being a **messaging standard** that enables disparate healthcare applications to exchange key sets of clinical and administrative data.

The HL7 standard specifications are unified by shared **reference models** of the healthcare and technical domains.

# Health Level Seven

The Health Level Seven International (HL7) is a not-for profit, ANSI-accredited **standards developing organization** dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of **health services**.

HL7 develops specifications, the most widely used being a **messaging standard** that enables disparate healthcare applications to exchange key sets of clinical and administrative data.

The HL7 standard specifications are unified by shared **reference models** of the healthcare and technical domains.

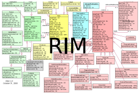
# Health Level Seven

The Health Level Seven International (HL7) is a not-for profit, ANSI-accredited **standards developing organization** dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of **health services**.

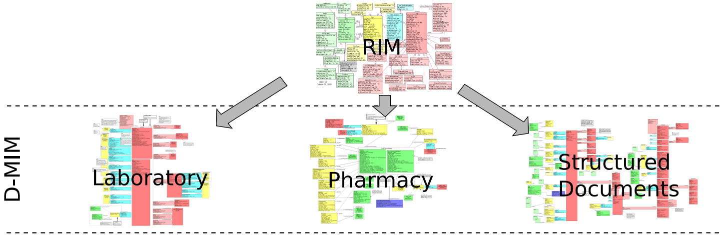
HL7 develops specifications, the most widely used being a **messaging standard** that enables disparate healthcare applications to exchange key sets of clinical and administrative data.

The HL7 standard specifications are unified by shared **reference models** of the healthcare and technical domains.

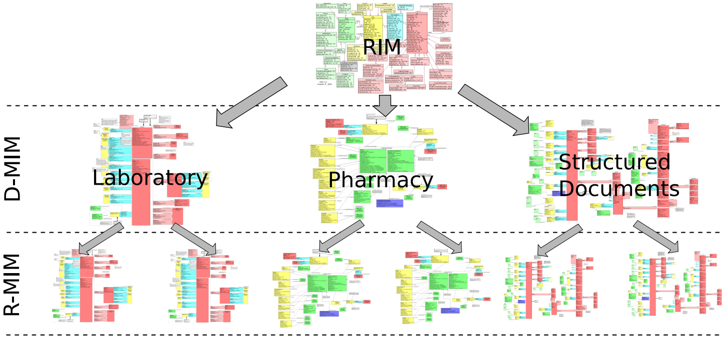
# Reference Models Overview



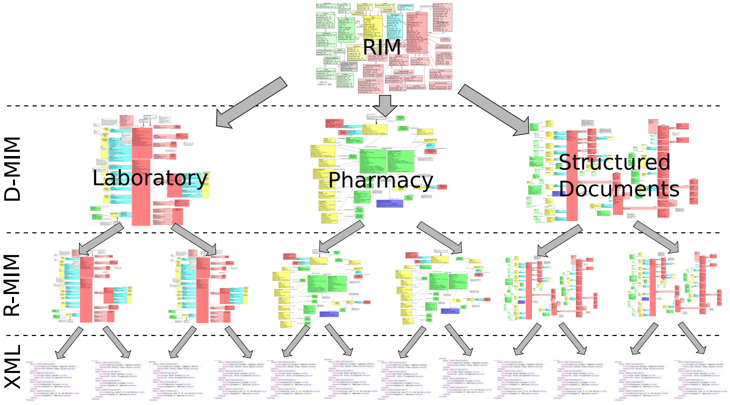
# Reference Models Overview



# Reference Models Overview



# Reference Models Overview



# Reference Models Overview
## Refinements

D-MIM models refine the RIM in three ways:

- The participants of one of the associations defined between RIM classes are refined in the subclasses.
- The multiplicities of an association defined between RIM classes are strengthened in the subclasses.
- The multiplicity of an attribute of a RIM class is strengthened in a subclass.

Note that it is not allowed to add new information.

R-MIM models refine D-MIM models in the same way.

# Reference Models Overview

## Refinements

D-MIM models refine the RIM in three ways:

- The participants of one of the associations defined between RIM classes are refined in the subclasses.
- The multiplicities of an association defined between RIM classes are strengthened in the subclasses.
- The multiplicity of an attribute of a RIM class is strengthened in a subclass.

Note that it is not allowed to add new information.

R-MIM models refine D-MIM models in the same way.

# Reference Models Overview

## Refinements

D-MIM models refine the RIM in three ways:

- The participants of one of the associations defined between RIM classes are refined in the subclasses.
- The multiplicities of an association defined between RIM classes are strengthened in the subclasses.
- The multiplicity of an attribute of a RIM class is strengthened in a subclass.

Note that it is not allowed to add new information.

R-MIM models refine D-MIM models in the same way.

# Reference Models Overview

## Refinements

D-MIM models refine the RIM in three ways:

- The participants of one of the associations defined between RIM classes are refined in the subclasses.
- The multiplicities of an association defined between RIM classes are strengthened in the subclasses.
- The multiplicity of an attribute of a RIM class is strengthened in a subclass.

Note that it is not allowed to add new information.

R-MIM models refine D-MIM models in the same way.

# Reference Models Overview
## Refinements

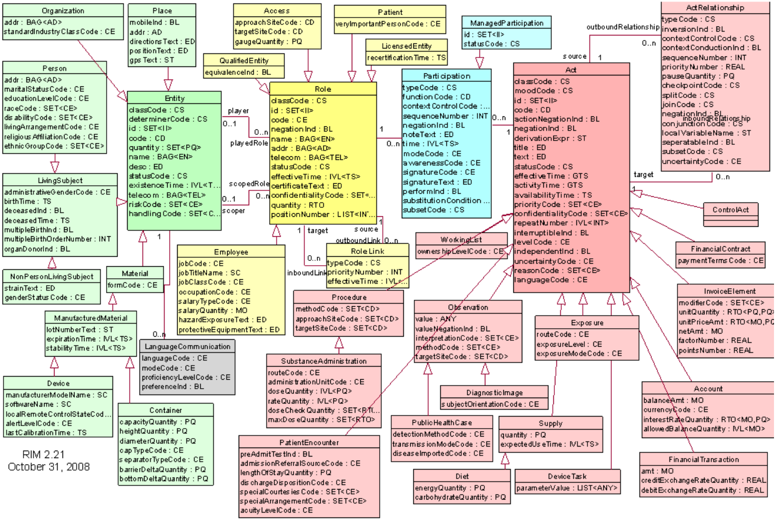
D-MIM models refine the RIM in three ways:

- The participants of one of the associations defined between RIM classes are refined in the subclasses.
- The multiplicities of an association defined between RIM classes are strengthened in the subclasses.
- The multiplicity of an attribute of a RIM class is strengthened in a subclass.

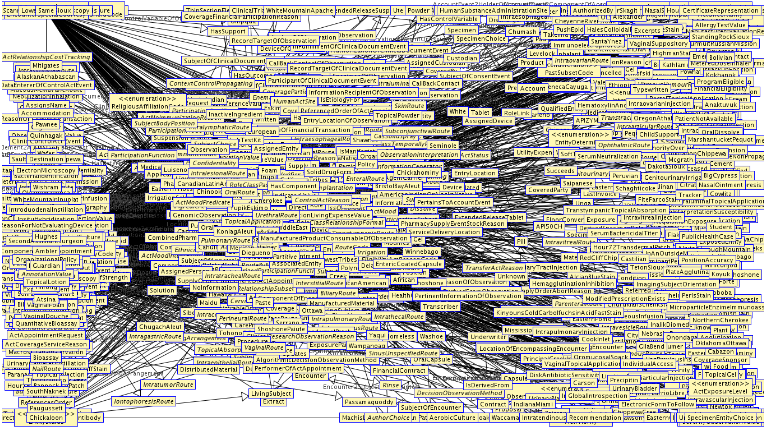
Note that it is not allowed to add new information.

R-MIM models refine D-MIM models in the same way.

# Reference Information Model (RIM)



# Reference Information Model (RIM)



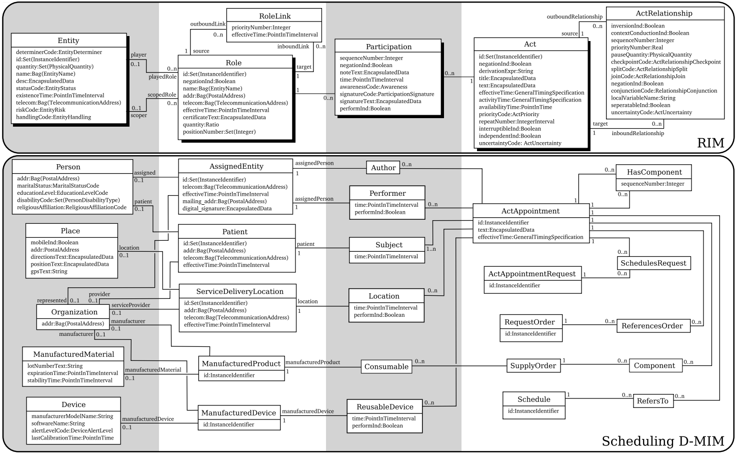
# Domain Message Information Model (D-MIM)

## HL7 Domains

- Administrative Management
- Account and Billing
- Claims & Reimbursement
- Patient Administration
- Materials Management
- Personnel Management
- Scheduling
- Blood Bank
- Care Provision
- Clinical Decision Support
- Clinical Document Architecture
- Clinical Genomics
- Diagnostic Imaging
- Laboratory
- Orders and Observations
- Medical Records
- Medication
- Pharmacy
- Public Health
- Regulated Products
- Regulated Studies
- Specimen Domain
- Therapeutic Devices

# Domain Message Information Model (D-MIM)

## Scheduling Domain



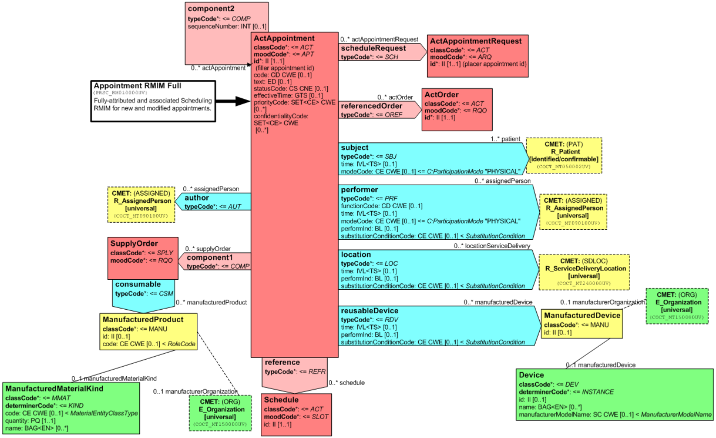
# Refined Message Information Model (R-MIM)

The R-MIM is a **subset of a D-MIM** that is used to express the information content for a message/document or set of messages/documents with annotations and **refinements** that are message/document **specific**.

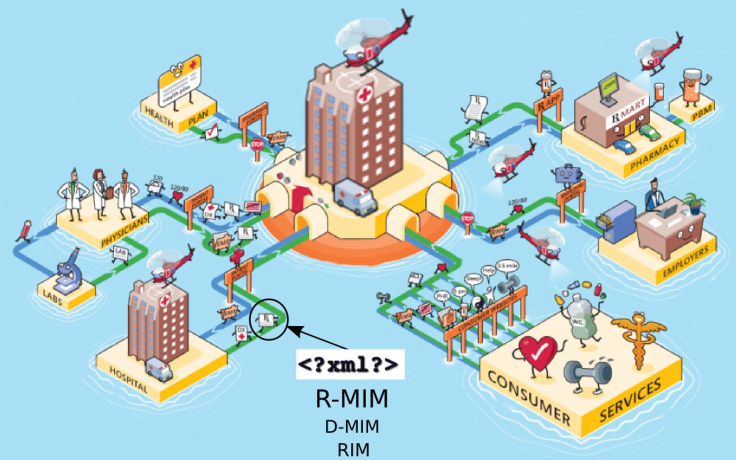
The content of an R-MIM is drawn from the D-MIM for the specific domain in which the R-MIM is used.

# Refined Message Information Model (R-MIM)

## Full Appointment R-MIM



# Interchanging Messages



# Outline

1. Introduction
   - IEEE 6th World Congress on Services (SERVICES 2010)
2. Health Level Seven
   - Healthcare Services
   - Reference Models Overview
   - RIM
   - D-MIM
   - R-MIM
3. Filtering HL7 Models
   - Overview
   - User Preferences
   - Filtering Measures
   - Interest Set
   - Filtered Information Model
4. Evaluation
   - Precision Analysis
   - Time Analysis
5. Conclusions

# Filtering HL7 Models

## Objective

Automatically provide a filtered information model of the whole HL7 models according to the user preferences.

## Filtered Information Model

A small information model that focus on the knowledge of the user's request.

Its reduced size and self-contained aspect make it easier to the user the comprehension and understandability of the focused knowledge.

# Filtering HL7 Models

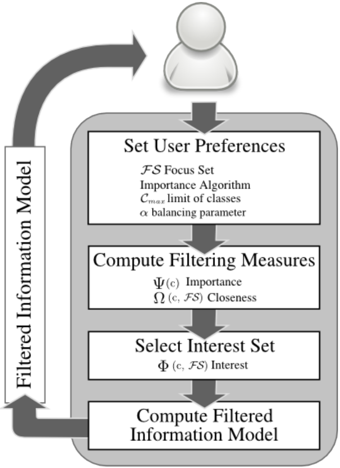
## Objective

Automatically provide a filtered information model of the whole HL7 models according to the user preferences.

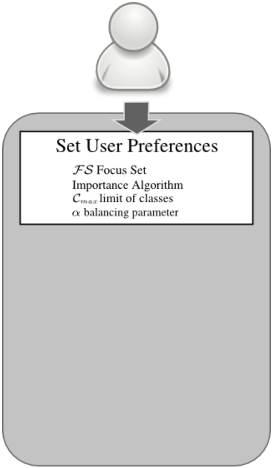
## Filtered Information Model

A small information model that focus on the knowledge of the user's request.
Its reduced size and self-contained aspect make it easier to the user the comprehension and understandability of the focused knowledge.

# Method Overview



# Method Overview



# Step 1: Setting the User Preferences

The user selects:

**Focus Set ($\mathcal{FS}$)**

a non-empty set of classes the user is interested in.

**Rejection Set ($\mathcal{RS}$)**

an optional set with those classes that have no interest to the user.

**Filter Size ($\mathcal{C}_{max}$)**

the amount of additional classes the user wants to obtain.

**Example**

$\mathcal{FS} = \{Patient, ActAppointment\}$ and $\mathcal{RS} = \emptyset$ and $\mathcal{C}_{max} = 12$

# Step 1: Setting the User Preferences

The user selects:

**Focus Set ($\mathcal{FS}$)**

a non-empty set of classes the user is interested in.

**Rejection Set ($\mathcal{RS}$)**

an optional set with those classes that have no interest to the user.

**Filter Size ($\mathcal{C}_{max}$)**

the amount of additional classes the user wants to obtain.

**Example**

$\mathcal{FS} = \{Patient, ActAppointment\}$ and $\mathcal{RS} = \emptyset$ and $\mathcal{C}_{max} = 12$

# Step 1: Setting the User Preferences

The user selects:

**Focus Set ($\mathcal{FS}$)**

a non-empty set of classes the user is interested in.

**Rejection Set ($\mathcal{RS}$)**

an optional set with those classes that have no interest to the user.

**Filter Size ($\mathcal{C}_{max}$)**

the amount of additional classes the user wants to obtain.

**Example**

$\mathcal{FS} = \{Patient, ActAppointment\}$ and $\mathcal{RS} = \emptyset$ and $\mathcal{C}_{max} = 12$

# Step 1: Setting the User Preferences

The user selects:

**Focus Set ($\mathcal{FS}$)**

a non-empty set of classes the user is interested in.

**Rejection Set ($\mathcal{RS}$)**

an optional set with those classes that have no interest to the user.

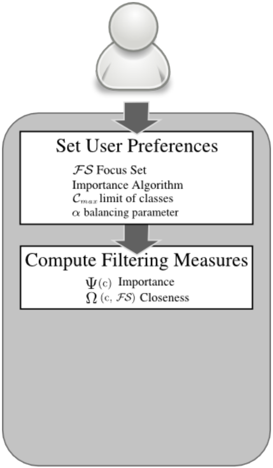
**Filter Size ($\mathcal{C}_{max}$)**

the amount of additional classes the user wants to obtain.

**Example**

$\mathcal{FS} = \{Patient, ActAppointment\}$ and $\mathcal{RS} = \emptyset$ and $\mathcal{C}_{max} = 12$

# Method Overview



# Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

## Importance ($\Psi$)

**Definition (Importance ($\Psi$))**

The importance $\Psi(c)$ of a class $c$ is a real number that measures the relative importance of that class in a model.

**Methods** [1]

- Occurrence Counting
- Link Analysis
- Instance-dependent

[1] *On Computing the Importance of Entity Types in Large Conceptual Schemas*. Villegas, A. and Olivé, A. *ER 2009*.

# Step 2: Compute Filtering Measures

## Importance ($\Psi$)

### Definition (Importance ($\Psi$))

The importance $\Psi(c)$ of a class $c$ is a real number that measures the relative importance of that class in a model.

### Methods [1]

- Occurrence Counting
- Link Analysis
- Instance-dependent

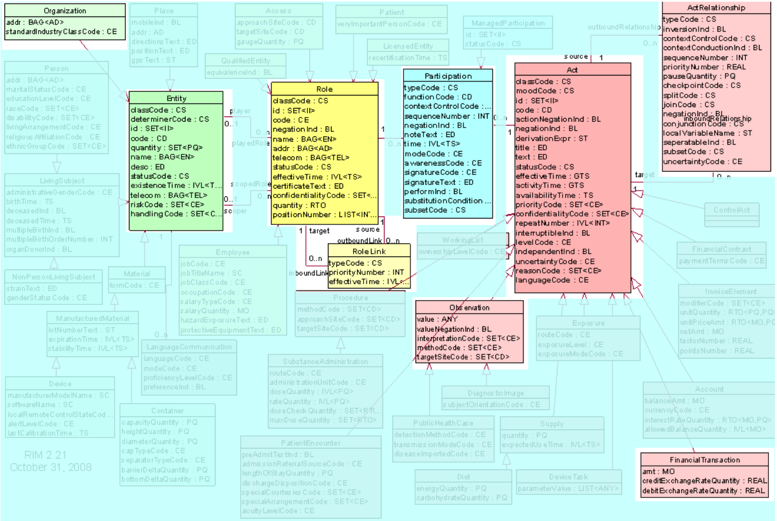
[1] *On Computing the Importance of Entity Types in Large Conceptual Schemas*. Villegas, A. and Olivé, A. *ER 2009*.

# Step 2: Compute Filtering Measures

## Importance ($\Psi$)

Top-10 Most Important Classes.

| Rank | Class | Importance $\Psi$ |
|---|---|---|
| 1 | Act | 7.51 |
| 2 | Role | 5.11 |
| 3 | ActRelationship | 4.03 |
| 4 | Participation | 3.67 |
| 5 | Entity | 3.5 |
| 6 | Observation | 2.64 |
| 7 | InfrastructureRoot | 1.81 |
| 8 | Organization | 1.72 |
| 9 | RoleLink | 1.59 |
| 10 | FinancialTransaction | 1.54 |


Organization
addr : BAG<AD>
standardIndustryClassCode : CE
Entity
classCode : CS
determinerCode : CS
id : SET<II>
code : CD
quantity : SET<PQ>
name : BAG<EN>
desc : ED
statusCode : CS
existenceTime : IVL<T...
telecom : BAG<TEL>
riskCode : SET<CE>
handlingCode : SET<C...
player
scoper
playedRole
scopedRole
Role
classCode : CS
id : SET<II>
code : CE
negationInd : BL
name : BAG<EN>
addr : BAG<AD>
telecom : BAG<TEL>
statusCode : CS
effectiveTime : IVL<TS>
certificateText : ED
confidentialityCode : SET<...
quantity : RTO
positionNumber : LIST<IN...
target
source
outboundLink
inboundLink
RoleLink
typeCode : CS
priorityNumber : INT
effectiveTime : IVL...
Participation
typeCode : CS
functionCode : CD
contextControlCode : ...
sequenceNumber : INT
negationInd : BL
noteText : ED
time : IVL<TS>
modeCode : CE
awarenessCode : CE
signatureCode : CE
signatureText : ED
performInd : BL
substitutionCondition ...
subsetCode : CS
Act
classCode : CS
moodCode : CS
id : SET<II>
code : CD
actionNegationInd : BL
negationInd : BL
derivationExpr : ST
title : ED
text : ED
statusCode : CS
effectiveTime : GTS
activityTime : GTS
availabilityTime : TS
priorityCode : SET<CE>
confidentialityCode : SET<CE>
repeatNumber : IVL<INT>
interruptibleInd : BL
levelCode : CE
independentInd : BL
uncertaintyCode : CE
reasonCode : SET<CE>
languageCode : CE
outboundRelationship
inboundRelationship
source
target
ActRelationship
typeCode : CS
inversionInd : BL
contextControlCode : CS
contextConductionInd : BL
sequenceNumber : INT
priorityNumber : REAL
pauseQuantity : PQ
checkpointCode : CS
splitCode : CS
joinCode : CS
negationInd : BL
conjunctionCode : CS
localVariableName : ST
seperatableInd : BL
subsetCode : CS
uncertaintyCode : CE
Observation
value : ANY
valueNegationInd : BL
interpretationCode : SET<CE>
methodCode : SET<CE>
targetSiteCode : SET<CD>
FinancialTransaction
amt : MO
creditExchangeRateQuantity : REAL
debitExchangeRateQuantity : REAL
Place
Access
Patient
ManagedParticipation
LicensedEntity
QualifiedEntity
Person
LivingSubject
NonPersonLivingSubject
Material
ManufacturedMaterial
Device
Container
Employee
LanguageCommunication
WorkingList
Procedure
SubstanceAdministration
PatientEncounter
ControlAct
FinancialContract
InvoiceElement
Exposure
DiagnosticImage
PublicHealthCase
Supply
Diet
DeviceTask
Account
RIM 2.21
October 31, 2008

## The importance problem

The importance of a class is an absolute metric that depends only on the whole set of HL7 models.

The metric is useful when a user wants to know which are the most important classes, but it is of little use when the user is interested in a specific subset of classes, independently from their importance.

What is needed then is a metric that measures the interest of a class with respect to the focus set.

## Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

## Closeness ($\Omega$)

There may be several ways to compute the closeness $\Omega(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to the classes of $\mathcal{FS}$.

Intuitively, the closeness of class $\mathrm{c}$ should be directly related to the inverse of the distance of $\mathrm{c}$ to the focus set $\mathcal{FS}$.

$$\Omega(\mathrm{c}, \mathcal{FS}) = \frac{|\mathcal{FS}|}{\sum\limits_{\mathrm{c}' \in \mathcal{FS}} \mathrm{d}(\mathrm{c}, \mathrm{c}')}$$

# Step 2: Compute Filtering Measures

## Closeness ($\Omega$)

There may be several ways to compute the closeness $\Omega(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to the classes of $\mathcal{FS}$.

Intuitively, the closeness of class $\mathrm{c}$ should be directly related to the inverse of the distance of $\mathrm{c}$ to the focus set $\mathcal{FS}$.

$$\Omega(\mathrm{c}, \mathcal{FS}) = \frac{|\mathcal{FS}|}{\sum_{\mathrm{c}' \in \mathcal{FS}} \mathrm{d}(\mathrm{c}, \mathrm{c}')}$$

# Step 2: Compute Filtering Measures

## Closeness ($\Omega$)

Intuitively, the closeness of class $\mathrm{c}$ should be directly related to the inverse of the distance of $\mathrm{c}$ to the focus set $\mathcal{FS}$.

$$\Omega(\mathrm{c}, \mathcal{FS}) = \frac{\boxed{|\mathcal{FS}|}}{\sum_{\mathrm{c}' \in \mathcal{FS}} \mathrm{d}(\mathrm{c}, \mathrm{c}')}$$

number of classes in $\mathcal{FS}$

# Step 2: Compute Filtering Measures

## Closeness ($\Omega$)

Intuitively, the closeness of class $\mathrm{c}$ should be directly related to the inverse of the distance of $\mathrm{c}$ to the focus set $\mathcal{FS}$.

$$\Omega(\mathrm{c}, \mathcal{FS}) = \frac{|\mathcal{FS}|}{\sum\limits_{\mathrm{c}' \in \mathcal{FS}} \mathrm{d}(\mathrm{c}, \mathrm{c}')}$$

minimum distance between $\mathrm{c}$ and $\mathrm{c}' \in \mathcal{FS}$

$\mathrm{c}$ and $\mathrm{c}'$ directly connected $\rightarrow \mathrm{d}(\mathrm{c}, \mathrm{c}') = 1$

otherwise $\rightarrow \mathrm{d}(\mathrm{c}, \mathrm{c}') =$ length of the shortest path between them

# Step 2: Compute Filtering Measures

## Closeness ($\Omega$)

Intuitively, the closeness of class $\mathrm{c}$ should be directly related to the inverse of the distance of $\mathrm{c}$ to the focus set $\mathcal{FS}$.

$$\Omega(\mathrm{c}, \mathcal{FS}) = \frac{|\mathcal{FS}|}{\sum\limits_{\mathrm{c}' \in \mathcal{FS}} \boxed{\mathrm{d}(\mathrm{c}, \mathrm{c}')}}$$

minimum distance between $\mathrm{c}$ and $\mathrm{c}' \in \mathcal{FS}$

$\mathrm{c}$ and $\mathrm{c}'$ directly connected $\rightarrow \mathrm{d}(\mathrm{c}, \mathrm{c}') = 1$

otherwise $\rightarrow \mathrm{d}(\mathrm{c}, \mathrm{c}') =$ length of the shortest path between them

# Step 2: Compute Filtering Measures

- **Importance** of classes ($\Psi$)
- **Closeness** between classes ($\Omega$)
- **Interest** of classes ($\Phi$)

# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

**Definition (Interest $\Psi$)**

The interest $\Phi(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to a focus set $\mathcal{FS}$ is a combination of the **importance** of $\mathrm{c}$ and its **closeness** to $\mathcal{FS}$.

$$\Phi(\mathrm{c}, \mathcal{FS}) = \alpha \times \Psi(\mathrm{c}) + (1 - \alpha) \times \Omega(\mathrm{c}, \mathcal{FS})$$

# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

**Definition (Interest $\Psi$)**

The interest $\Phi(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to a focus set $\mathcal{FS}$ is a combination of the **importance** of $\mathrm{c}$ and its **closeness** to $\mathcal{FS}$.

$$\Phi(\mathrm{c}, \mathcal{FS}) = \alpha \times \Psi(\mathrm{c}) + (1 - \alpha) \times \Omega(\mathrm{c}, \mathcal{FS})$$

# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

**Definition (Interest $\Psi$)**

The interest $\Phi(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to a focus set $\mathcal{FS}$ is a combination of the **importance** of $\mathrm{c}$ and its **closeness** to $\mathcal{FS}$.

$$\Phi(\mathrm{c}, \mathcal{FS}) = \boxed{\alpha \times \Psi(\mathrm{c})} + (1 - \alpha) \times \Omega(\mathrm{c}, \mathcal{FS})$$

Component of Importance

# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

**Definition (Interest $\Psi$)**

The interest $\Phi(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to a focus set $\mathcal{FS}$ is a combination of the **importance** of $\mathrm{c}$ and its **closeness** to $\mathcal{FS}$.

$$\Phi(\mathrm{c}, \mathcal{FS}) = \alpha \times \Psi(\mathrm{c}) + (1-\alpha) \times \Omega(\mathrm{c}, \mathcal{FS})$$

Component of Closeness

# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

**Definition (Interest $\Psi$)**

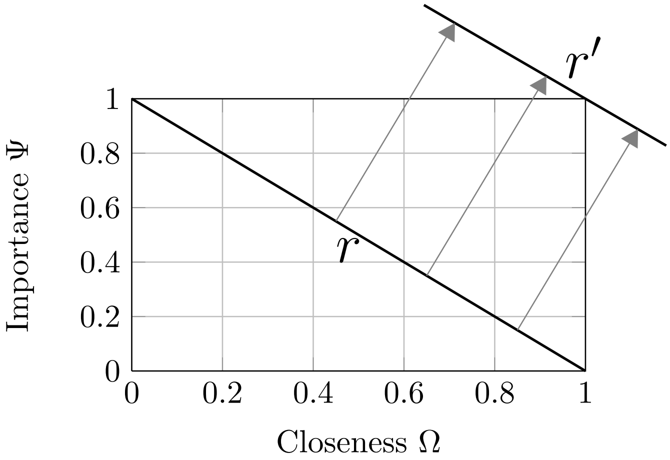
The interest $\Phi(\mathrm{c}, \mathcal{FS})$ of a class $\mathrm{c}$ with respect to a focus set $\mathcal{FS}$ is a combination of the **importance** of $\mathrm{c}$ and its **closeness** to $\mathcal{FS}$.

$$\Phi(\mathrm{c}, \mathcal{FS}) = \alpha \times \Psi(\mathrm{c}) + (1 - \alpha) \times \Omega(\mathrm{c}, \mathcal{FS})$$

Balancing Parameter (default $\alpha = 0.5$)

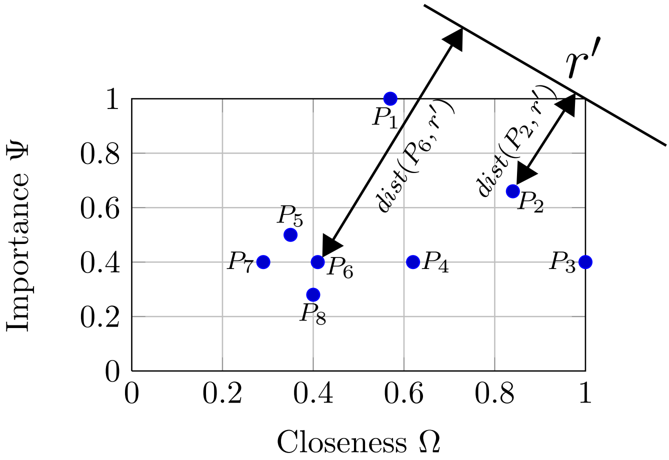
# Step 2: Compute Filtering Measures

## Interest ($\Phi$)

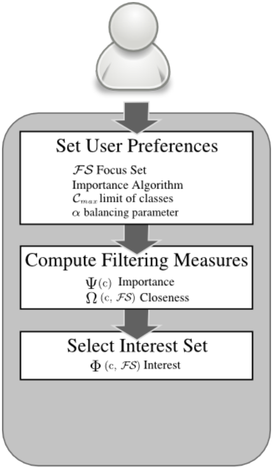


# Step 2: Compute Filtering Measures

## Interest ($\Phi$)



# Method Overview



# Step 3: Select Interest Set

Select the top classes of the ranking produced by the computation of the interest $\Phi$ s.t. $|\text{Interest Set}| = \mathcal{C}_{\max} - |\mathcal{FS}|$.

Most Interesting classes with regard to $\mathcal{FS} = \{\textit{Patient, ActAppointment}\}$.

| Rank | Class ($c$) | Importance $\Psi(c)$ | Distance $d(c, \textit{Patient})$ | Distance $d(c, \textit{ActAppointment})$ | Closeness $\Omega(c, \mathcal{FS})$ | Interest $\Phi(c, \mathcal{FS})$ |
|---|---|---|---|---|---|---|
| 1 | Organization | 1.72 | 1 | 3 | 0.5 | 1.11 |
| 2 | Person | 1.22 | 1 | 3 | 0.5 | 0.86 |
| 3 | ServiceDeliveryLocation | 0.79 | 2 | 2 | 0.5 | 0.65 |
| 4 | AssignedPerson | 0.72 | 2 | 2 | 0.5 | 0.61 |
| 5 | SubjectOfActAppointment | 0.11 | 1 | 1 | 1.0 | 0.56 |
| 6 | ManufacturedDevice | 0.55 | 2 | 2 | 0.5 | 0.53 |
| 7 | LocationOfActAppointment | 0.26 | 3 | 1 | 0.5 | 0.38 |
| 8 | ReusableDeviceOfActAppointment | 0.19 | 3 | 1 | 0.5 | 0.35 |
| 9 | SubjectOfAccountEvent | 0.13 | 1 | 3 | 0.5 | 0.32 |
| 10 | AuthorOfActAppointment | 0.12 | 3 | 1 | 0.5 | 0.31 |

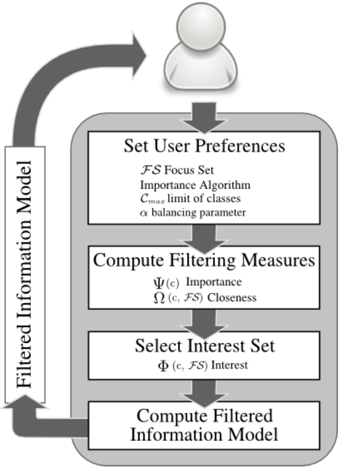
# Step 3: Select Interest Set

Select the top classes of the ranking produced by the computation of the interest $\Phi$ s.t. $|\text{Interest Set}| = \mathcal{C}_{\max} - |\mathcal{FS}|$.

Most Interesting classes with regard to $\mathcal{FS} = \{$*Patient*, *ActAppointment*$\}$.

| Rank | Class ($c$) | Importance $\Psi(\mathrm{c})$ | Distance $\mathrm{d}(\mathrm{c},$ *Patient*$)$ | Distance $\mathrm{d}(\mathrm{c},$ *ActAppointment*$)$ | Closeness $\Omega(\mathrm{c}, \mathcal{FS})$ | Interest $\Phi(c, \mathcal{FS})$ |
|---|---|---|---|---|---|---|
| 1 | Organization | 1.72 | 1 | 3 | 0.5 | 1.11 |
| 2 | Person | 1.22 | 1 | 3 | 0.5 | 0.86 |
| 3 | ServiceDeliveryLocation | 0.79 | 2 | 2 | 0.5 | 0.65 |
| 4 | AssignedPerson | 0.72 | 2 | 2 | 0.5 | 0.61 |
| 5 | SubjectOfActAppointment | 0.11 | 1 | 1 | 1.0 | 0.56 |
| 6 | ManufacturedDevice | 0.55 | 2 | 2 | 0.5 | 0.53 |
| 7 | LocationOfActAppointment | 0.26 | 3 | 1 | 0.5 | 0.38 |
| 8 | ReusableDeviceOfActAppointment | 0.19 | 3 | 1 | 0.5 | 0.35 |
| 9 | SubjectOfAccountEvent | 0.13 | 1 | 3 | 0.5 | 0.32 |
| 10 | AuthorOfActAppointment | 0.12 | 3 | 1 | 0.5 | 0.31 |

# Method Overview



# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

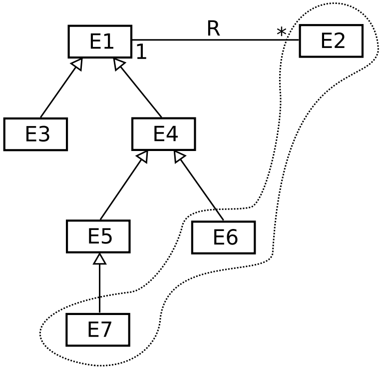
# Step 4: Compute Filtered Information Model

## Filtered Information Model (FIM) Construction

- **Classes**
  - $\mathcal{FS}$ Classes
  - Interest Set Classes
  - Auxiliary Classes
- **Associations**
  - Participant Classes are in FIM
  - Participant Classes are superclasses of classes in FIM
    - Project the association to subclasses
- **Generalization-Specialization Relationships**
  - Superclass and subclass are in FIM
  - Indirect path of generalizations induces superclass and subclass in FIM
    - Mark generalization as indirect

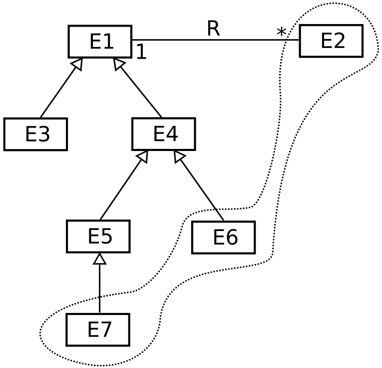
# Step 4: Compute Filtered Information Model

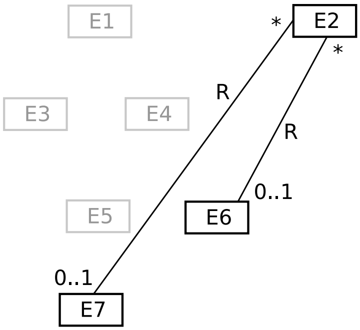
## Projection of Association



# Step 4: Compute Filtered Information Model

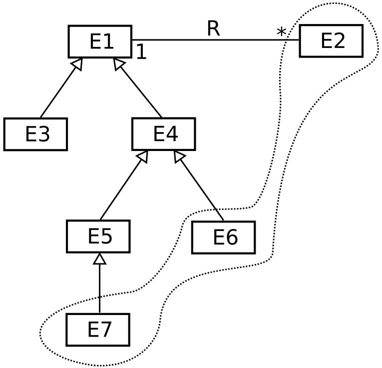
## Projection of Association

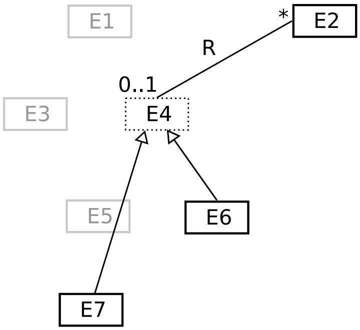


# Step 4: Compute Filtered Information Model

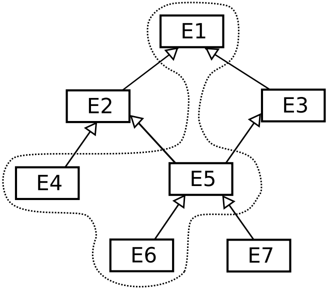
## Projection of Association



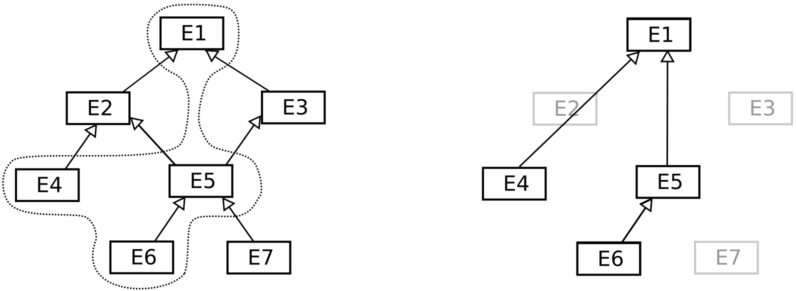
# Step 4: Compute Filtered Information Model

## Generalization-Specialization Relationships



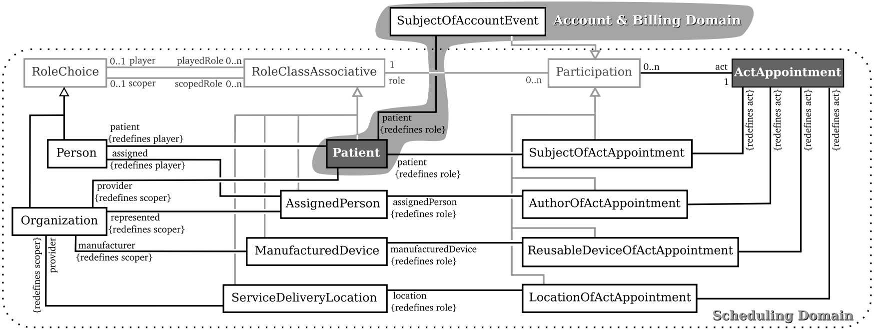
# Step 4: Compute Filtered Information Model

## Generalization-Specialization Relationships



# Step 4: Compute Filtered Information Model

$\mathcal{FS}$ = {Patient, ActAppointment} and $\mathcal{C}_{max} = 12$.



# Outline

1. Introduction
   - IEEE 6th World Congress on Services (SERVICES 2010)
2. Health Level Seven
   - Healthcare Services
   - Reference Models Overview
   - RIM
   - D-MIM
   - R-MIM
3. Filtering HL7 Models
   - Overview
   - User Preferences
   - Filtering Measures
   - Interest Set
   - Filtered Information Model
4. Evaluation
   - Precision Analysis
   - Time Analysis
5. Conclusions

# Evaluation

To find a measure that reflects the ability of our method to satisfy the user is a complicated task.

However, there exists measurable quantities in the field of information retrieval that can be applied to our context:

- The ability of the method to withhold non-relevant knowledge (*precision*)
- The interval between the request being made and the answer being given (*time*)

# Evaluation

To find a measure that reflects the ability of our method to satisfy the user is a complicated task.

However, there exists measurable quantities in the field of information retrieval that can be applied to our context:

- The ability of the method to withhold non-relevant knowledge (*precision*)
- The interval between the request being made and the answer being given (*time*)

# Evaluation

To find a measure that reflects the ability of our method to satisfy the user is a complicated task.

However, there exists measurable quantities in the field of information retrieval that can be applied to our context:

- The ability of the method to withhold non-relevant knowledge (*precision*)
- The interval between the request being made and the answer being given (*time*)

# Evaluation

To find a measure that reflects the ability of our method to satisfy the user is a complicated task.

However, there exists measurable quantities in the field of information retrieval that can be applied to our context:

- The ability of the method to withhold non-relevant knowledge (*precision*)
- The interval between the request being made and the answer being given (*time*)

# Precision Analysis

The precision of a method is defined as the percentage of relevant knowledge presented to the user.

We use the concept of precision applied to HL7 universal domains (specified with D-MIM's).

$$Precision = \frac{|\{\text{relevant classes}\}| \ \cap \ |\{\text{retrieved classes}\}|}{|\{\text{retrieved classes}\}|}$$

# Precision Analysis

The precision of a method is defined as the percentage of relevant knowledge presented to the user.

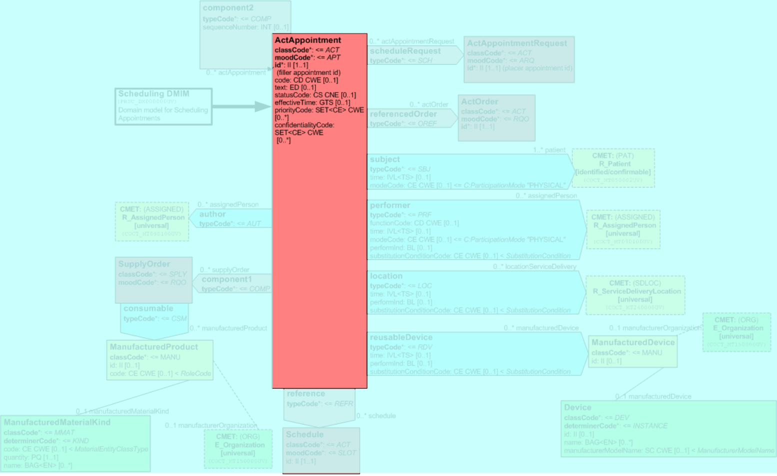
We use the concept of precision applied to HL7 universal domains (specified with D-MIM's).

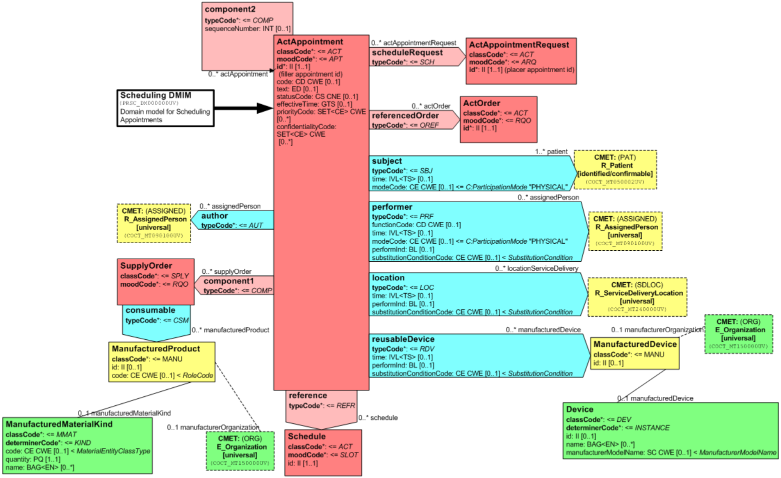
$$Precision = \frac{|\{\text{relevant classes}\}| \ \cap \ |\{\text{retrieved classes}\}|}{|\{\text{retrieved classes}\}|}$$

# Precision Analysis

Each domain contains a main class which is the central point of knowledge to the users interested in such domain. The other classes presented in the domain conform the relevant knowledge related to the main class.

A common situation for a user is to focus on the main class of a domain and to navigate through the D-MIM to understand its related knowledge.


component2
typeCode*: <= COMP
sequenceNumber: INT [0..1]
0..* actAppointment
ActAppointment
classCode*: <= ACT
moodCode*: <= APT
id*: II [1..1]
(filler appointment id)
code: CD CWE [0..1]
text: ED [0..1]
statusCode: CS CNE [0..1]
effectiveTime: GTS [0..1]
priorityCode: SET<CE> CWE [0..*]
confidentialityCode: SET<CE> CWE [0..*]
Scheduling DMIM
Domain model for Scheduling Appointments
0..* actAppointmentRequest
scheduleRequest
typeCode*: <= SCH
ActAppointmentRequest
classCode*: <= ACT
moodCode*: <= ARQ
id*: II [1..1] (placer appointment id)
0..* actOrder
referencedOrder
typeCode*: <= OREF
ActOrder
classCode*: <= ACT
moodCode*: <= RQO
id*: II [1..1]
1..* patient
subject
typeCode*: <= SBJ
time: IVL<TS> [0..1]
modeCode: CE CWE [0..1] <= C:ParticipationMode "PHYSICAL"
CMET: (PAT)
R_Patient
[identified/confirmable]
0..* assignedPerson
performer
typeCode*: <= PRF
functionCode: CD CWE [0..1]
time: IVL<TS> [0..1]
modeCode: CE CWE [0..1] <= C:ParticipationMode "PHYSICAL"
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
CMET: (ASSIGNED)
R_AssignedPerson
[universal]
author
typeCode*: <= AUT
SupplyOrder
classCode*: <= SPLY
moodCode*: <= RQO
0..* supplyOrder
component1
typeCode*: <= COMP
consumable
typeCode*: <= CSM
0..* manufacturedProduct
0..* locationServiceDelivery
location
typeCode*: <= LOC
time: IVL<TS> [0..1]
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
CMET: (SDLOC)
R_ServiceDeliveryLocation
[universal]
0..* manufacturedDevice
reusableDevice
typeCode*: <= RDV
time: IVL<TS> [0..1]
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
0..1 manufacturerOrganization
ManufacturedDevice
classCode*: <= MANU
id: II [0..1]
CMET: (ORG)
E_Organization
[universal]
ManufacturedProduct
classCode*: <= MANU
id: II [0..1]
code: CE CWE [0..1] < RoleCode
0..1 manufacturedMaterialKind
ManufacturedMaterialKind
classCode*: <= MMAT
determinerCode*: <= KIND
code: CE CWE [0..1] < MaterialEntityClassType
quantity: PQ [1..1]
name: BAG<EN> [0..*]
0..1 manufacturerOrganization
reference
typeCode*: <= REFR
0..* schedule
Schedule
classCode*: <= ACT
moodCode*: <= SLOT
id: II [1..1]
0..1 manufacturedDevice
Device
classCode*: <= DEV
determinerCode*: <= INSTANCE
id: II [0..1]
name: BAG<EN> [0..*]
manufacturerModelName: SC CWE [0..1] < ManufacturerModelName


component2
typeCode*: <= COMP
sequenceNumber: INT [0..1]
0..* actAppointment
ActAppointment
classCode*: <= ACT
moodCode*: <= APT
id*: II [1..1]
(filler appointment id)
code: CD CWE [0..1]
text: ED [0..1]
statusCode: CS CNE [0..1]
effectiveTime: GTS [0..1]
priorityCode: SET<CE> CWE [0..*]
confidentialityCode: SET<CE> CWE [0..*]
0..* actAppointmentRequest
scheduleRequest
typeCode*: <= SCH
ActAppointmentRequest
classCode*: <= ACT
moodCode*: <= ARQ
id*: II [1..1] (placer appointment id)
Scheduling DMIM
(PRSC_DM000000UV)
Domain model for Scheduling Appointments
0..* actOrder
referencedOrder
typeCode*: <= OREF
ActOrder
classCode*: <= ACT
moodCode*: <= RQO
id*: II [1..1]
1..* patient
subject
typeCode*: <= SBJ
time: IVL<TS> [0..1]
modeCode: CE CWE [0..1] <= C:ParticipationMode "PHYSICAL"
CMET: (PAT)
R_Patient
[identified/confirmable]
(COCT_MT050002UV)
0..* assignedPerson
author
typeCode*: <= AUT
CMET: (ASSIGNED)
R_AssignedPerson
[universal]
(COCT_MT090100UV)
0..* assignedPerson
performer
typeCode*: <= PRF
functionCode: CD CWE [0..1]
time: IVL<TS> [0..1]
modeCode: CE CWE [0..1] <= C:ParticipationMode "PHYSICAL"
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
CMET: (ASSIGNED)
R_AssignedPerson
[universal]
(COCT_MT090100UV)
SupplyOrder
classCode*: <= SPLY
moodCode*: <= RQO
0..* supplyOrder
component1
typeCode*: <= COMP
0..* locationServiceDelivery
location
typeCode*: <= LOC
time: IVL<TS> [0..1]
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
CMET: (SDLOC)
R_ServiceDeliveryLocation
[universal]
(COCT_MT240000UV)
consumable
typeCode*: <= CSM
0..* manufacturedProduct
ManufacturedProduct
classCode*: <= MANU
id: II [0..1]
code: CE CWE [0..1] < RoleCode
0..* manufacturedDevice
reusableDevice
typeCode*: <= RDV
time: IVL<TS> [0..1]
performInd: BL [0..1]
substitutionConditionCode: CE CWE [0..1] < SubstitutionCondition
0..1 manufacturerOrganization
ManufacturedDevice
classCode*: <= MANU
id: II [0..1]
CMET: (ORG)
E_Organization
[universal]
(COCT_MT150000UV)
reference
typeCode*: <= REFR
0..* schedule
0..1 manufacturedMaterialKind
ManufacturedMaterialKind
classCode*: <= MMAT
determinerCode*: <= KIND
code: CE CWE [0..1] < MaterialEntityClassType
quantity: PQ [1..1]
name: BAG<EN> [0..*]
0..1 manufacturerOrganization
CMET: (ORG)
E_Organization
[universal]
(COCT_MT150000UV)
Schedule
classCode*: <= ACT
moodCode*: <= SLOT
id: II [1..1]
0..1 manufacturedDevice
Device
classCode*: <= DEV
determinerCode*: <= INSTANCE
id: II [0..1]
name: BAG<EN> [0..*]
manufacturerModelName: SC CWE [0..1] < ManufacturerModelName

# Precision Analysis

We simulate the generation of a D-MIM from its main class.

## Initialization

- $\mathcal{FS}$ = main class of the D-MIM
- $\mathcal{C}_{max}$ = number of classes of the D-MIM

This way, we will obtain a filtered information model with the same number of classes as such domain.

## Following Iterations

- $\mathcal{FS}$ = main class of the D-MIM
- $\mathcal{C}_{max}$ = number of classes of the D-MIM
- $\mathcal{RS}$ = includes non-relevant classes retrieved in the previous iteration

# Precision Analysis

We simulate the generation of a D-MIM from its main class.

## Initialization

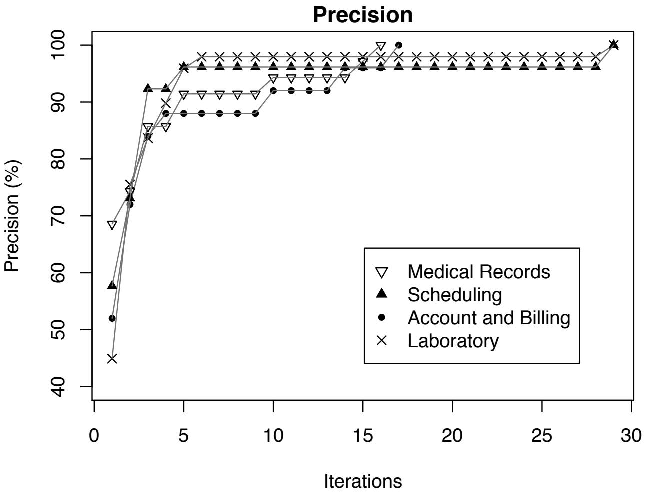
- $\mathcal{FS}$ = main class of the D-MIM
- $\mathcal{C}_{max}$ = number of classes of the D-MIM

This way, we will obtain a filtered information model with the same number of classes as such domain.

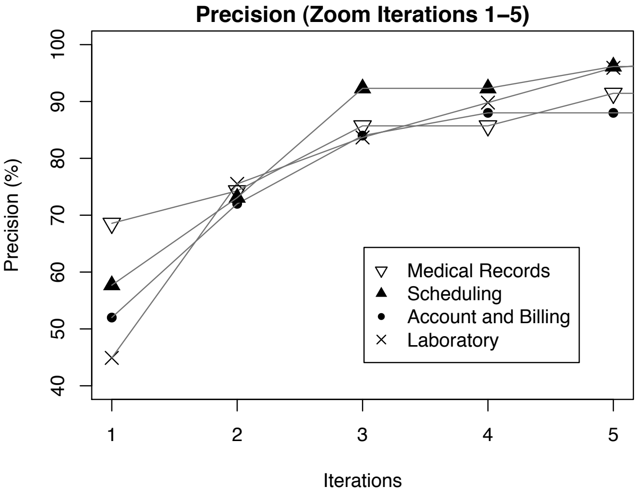
## Following Iterations

- $\mathcal{FS}$ = main class of the D-MIM
- $\mathcal{C}_{max}$ = number of classes of the D-MIM
- $\mathcal{RS}$ = includes non-relevant classes retrieved in the previous iteration

# Precision Analysis



# Precision Analysis



# Precision Analysis

The test reveals that to reach more than 80% of the relevant classes of a domain, only three iterations are required.

# Time Analysis

A good method does not only require precision, but it also needs to present the results in an acceptable time according to the user.

**Test**

Record the time lapse between the request of knowledge, i.e. once a focus set $\mathcal{FS}$ has been indicated by the user, and the receipt of the filtered information model.

# Time Analysis

A good method does not only require precision, but it also needs to present the results in an acceptable time according to the user.

## Test

Record the time lapse between the request of knowledge, i.e. once a focus set $\mathcal{FS}$ has been indicated by the user, and the receipt of the filtered information model.

# Time Analysis

It is expected that as we increase the size of the focus set, the time will increase linearly.

**Reason**

Our method computes the distances from each class in the focus set to all the rest of classes. This computation requires the same time (in average) for each class in the focus set.

Therefore, the more classes we have in a focus set, the more the time our method spends in computing distances.

# Time Analysis

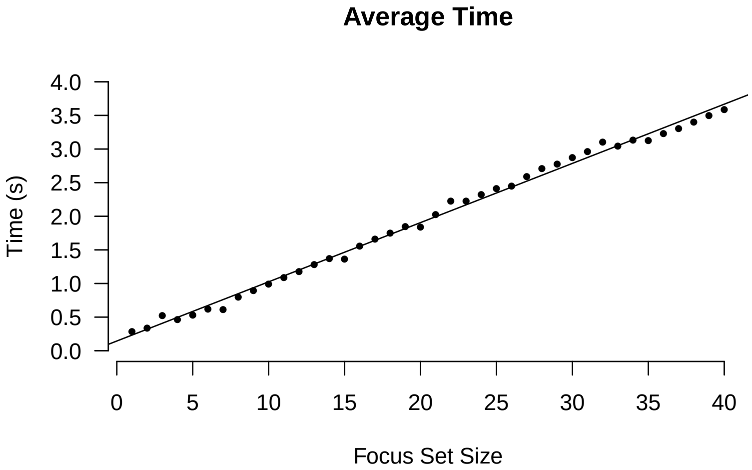
It is expected that as we increase the size of the focus set, the time will increase linearly.

**Reason**

Our method computes the distances from each class in the focus set to all the rest of classes. This computation requires the same time (in average) for each class in the focus set.

Therefore, the more classes we have in a focus set, the more the time our method spends in computing distances.

# Time Analysis



# Time Analysis

According to the expected use of our method, having a focus set $\mathcal{FS}$ of 40 classes is not a common situation (although possible).

Sizes of focus set up to 10 classes are more realistic, in which case the average time does not exceed one second.

# Time Analysis

According to the expected use of our method, having a focus set $\mathcal{FS}$ of 40 classes is not a common situation (although possible).

Sizes of focus set up to 10 classes are more realistic, in which case the average time does not exceed one second.

# Outline

1 Introduction
- IEEE 6th World Congress on Services (SERVICES 2010)

2 Health Level Seven
- Healthcare Services
- Reference Models Overview
- RIM
- D-MIM
- R-MIM

3 Filtering HL7 Models
- Overview
- User Preferences
- Filtering Measures
- Interest Set
- Filtered Information Model

4 Evaluation
- Precision Analysis
- Time Analysis

5 Conclusions

# Conclusions

**HL7 information models are very large**. The wealth of knowledge they contain makes them **very useful** to their potential target audience.

However, the size and the organization of these models make it **difficult to manually extract knowledge** from them. This task is basic for the improvement of services provided by HL7 affiliates, vendors and other organizations that use those models for the development of health systems.

What is needed is a **tool that improves the usability** of HL7 models for that task. Ours selects the most interesting classes from those models, including their defined knowledge in the original models.

The experiments show that **our prototype tool recovers** more than **80%** of the knowledge of a D-MIM in **three iterations**, with an **average time** per iteration that for common uses does not exceed **one second**.

# Improving the Usability of HL7 Information Models by Automatic Filtering

Antonio Villegas[1] Antoni Olivé[1] Josep Vilalta[2]

[1]Services and Information Systems Engineering Department
Universitat Politècnica de Catalunya

[2]HL7 Education & e-Learning Services
HL7 Spain (Health Level Seven International)

ESSI Seminar
June 2, 2010