



KDI ● **Knowledge and Data Integration**

Evaluation

iTelos Inception & Informal Modeling Phase

Fausto Giunchiglia

Contents

1 Evaluation on Inception phase

2 Evaluation on Informal Modeling phase

Contents

1 Evaluation on Inception phase

2 Evaluation on Informal Modeling phase

Evaluation purpose on Inception phase

In **schema level**, we have a set of Competency queries (CQs) and several collected ontologies.

We aim to measure:

- If the collect ontologies cover CQs, using metric **coverage**.
- If the collect ontologies bring additional information to CQs, using metric **Extensiveness**.

Examples: CQ vs Ont

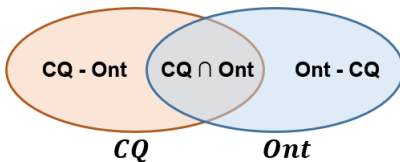
Given a set of Competency Query (CQ) , the coverage (Cov) of the aligned ontology (Ont) is:

Etype Coverage
$$Cov(CQ_c) = \frac{|CQ_c \cap Ont_c|}{|CQ_c|}$$

$Cov = 1$ Full coverage
 $0.6 < Cov < 0.8$ Ideal
 $Cov = 0$ No coverage

Property Coverage
$$Cov(CQ_p) = \frac{|CQ_p \cap Ont_p|}{|CQ_p|}$$

$Cov = 1$ Full coverage
 $0.6 < Cov < 0.8$ Ideal
 $Cov = 0$ No coverage



Examples: CQ vs Ont

Given a set of Competency Query (CQ) , the Extensiveness (Ext) of the aligned ontology (Ont) is:

$$\text{Etype Extensiveness } Ext(CQ_c) = \frac{|Ont_c - CQ_c|}{|CQ_c| + |Ont_c|}$$

$Ext = 1$ Full Extensiveness

$Ext \approx 0.5$ Ideal

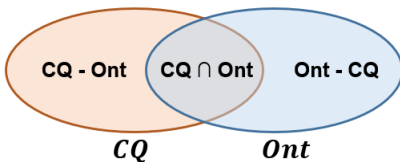
$Ext = 0$ No Extensiveness

$$\text{Property Extensiveness } Ext(CQ_p) = \frac{|Ont_p - CQ_p|}{|CQ_p| + |Ont_p|}$$

$Ext = 1$ Full Extensiveness

$Ext \approx 0.5$ Ideal

$Ext = 0$ No Extensiveness



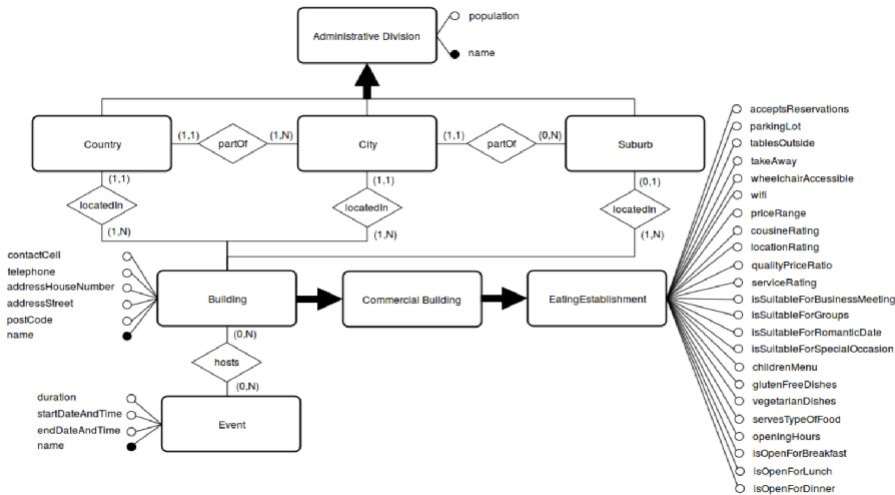
Examples: CQs

- List all the **places** where one can eat which have all **rating** higher than 4 stars.
- List all the **eating establishments** which are **suitable for groups** and also provide **children menu**.
- Give me **facilities** with **vegetarians** menu and **price** range medium-low.
- Give me the contacts of all the **eating establishments** that accept **reservations** and have a **parking lot**.
- Where to eat **pizza** for **lunch**?

Classes in CQs: $C_c = \{\text{place, eatingEstablishment, facilities, pizza}\}$
(Num class = 4)

Properties in CQs: $C_p = \{\text{rating, suitableForGroup, childrenMenu, Vegetarians, contact, reservation, parkingLot, lunch}\}$
(Num property = 8)

Examples: Reference Ontologies



Examples: Reference Ontologies

Classes/Etypes in ontology:

$C_c = \{\text{AdministrativeDivision, Country, City, Suburb, Building, CommercialBuilding, EatingEstablishment, Event}\}$
(Num class = 8)

Properties in ontology:

$C_p = \{\text{rating, suitableForGroup, childrenMenu, Vegetarians, contact, reservation, parkingLot, OpenForlunch}\}$
(Num property = 38)

Examples: CQ vs Ont

Given the example CQ , and the example reference ontology Ont , we have:

$$\text{Etype Coverage } Cov(CQ_c) = \frac{|CQ_c \cap Ont_c| (2)}{|CQ_c| (4)} = 0.5$$

$$\text{Etype Extensiveness } Ext(CQ_c) = \frac{|Ont_c - CQ_c| (6)}{|CQ_c| + |Ont_c| (8 + 4)} = 0.5$$

$$\text{Property Coverage } Cov(CQ_p) = \frac{|CQ_p \cap Ont_p| (6)}{|CQ_p| (8)} = 0.75$$

$$\text{Property Extensiveness } Ext(CQ_p) = \frac{|Ont_p - CQ_p| (6)}{|CQ_p| + |Ont_p| (38 + 8)} = 0.13$$

Notice that:

- Intersection information (coverage) should more likely belongs to common or core category.
- Additional information (extensiveness) should be core and contextual information.

Evaluation purpose on Inception phase

In **data level**, we have a set of CQs and several collected datasets/schema.

We aim to measure:

- If the collect datasets cover CQs, using metric **coverage**.
- If the collect datasets are much different from CQs, using metric **Sparsity**.

Examples: CQ vs Dataset

Given a set of Competency Query (CQ) , the coverage (Cov) of the aligned dataset (D) is:

Etype
Coverage
$$Cov(CQ_c) = \frac{|CQ_c \cap D_c|}{|CQ_c|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

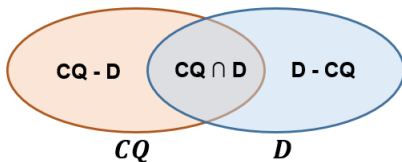
$Cov = 0$ No coverage

Property
Coverage
$$Cov(CQ_p) = \frac{|CQ_p \cap D_p|}{|CQ_p|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

$Cov = 0$ No coverage



Examples: CQ vs Dataset

Given a set of Competency Query (CQ), the sparsity (*Spr*) of the aligned dataset (*D*) is:

Etype sparsity
$$Spr(CQ_c) = \frac{|CQ_c - D_c| + |D_c - CQ_c|}{|CQ_c| + |D_c|} = 1 - 2 * \frac{|CQ_c \cap D_c|}{|CQ_c| + |D_c|}$$

$Spr = 1$ Full Sparsity

$Spr \approx 0.5$ Ideal

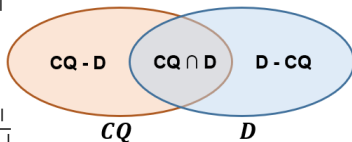
$Spr = 0$ No Sparsity

Property sparsity
$$Spr(CQ_p) = \frac{|CQ_p - D_p| + |D_p - CQ_p|}{|CQ_p| + |D_p|} = 1 - 2 * \frac{|CQ_p \cap D_p|}{|CQ_p| + |D_p|}$$

$Spr = 1$ Full Sparsity

$Spr \approx 0.5$ Ideal

$Spr = 0$ No Sparsity



Examples Dataset

Classes in dataset: (Num = 11), **Properties in dataset:** (Num = 40)

Establishment Type

- ☒ Restaurants
- ☐ Dessert
- ☐ Coffee & Tea
- More ▾

Reservations

- ☐ Online Reservations
- ☐ Restaurant Deals
- ☐ Available Tonight

Cuisines & Dishes

- ☐ Italian
- ☒ Pizza
- ☐ Mediterranean
- More ▾

Dietary Restrictions

- ☐ Vegetarian Friendly
- ☐ Vegan Options
- ☐ Gluten Free Options

Meals

- ☐ Breakfast
- ☐ Brunch
- ☐ Lunch
- ☐ Dinner

Price

- ☐ Cheap Eats



Oro Stube Povo

404 reviews

#2 of 101 Results

\$ - \$\$\$, Italian, Pizza, Mediterranean, Vegetarian Friendly, Vegan Options, Glu...

"Avant-garde Italian food" 10/01/2018
"nice" 12/21/2016



Pizzeria da Albert

1,666 reviews

#3 of 101 Results

\$ - \$\$\$, Italian, Pizza, Vegetarian Friendly, Vegan Options

"Sensational pizza!" 06/16/2017
"... euros for margherita and the most e..." 07/19/2017



Ristorante Pizzeria Al Duomo

230 reviews

#4 of 101 Results

\$ - \$\$\$, Italian, Pizza, Mediterranean, European, Vegetarian Friendly, Vegan O...

"Great pizza in the heart of Trento" 06/03/2018
"Could Not Ask For More" 07/17/2018



Uva e Menta

1,091 reviews

#5 of 101 Results

\$ - \$\$\$, Italian, Brew Pub, Pizza, Mediterranean, Vegetarian Friendly, Vegan O...

"Amazing pizzas, friendly staff, service a..." 07/31/2018
"Good Brew, Tasty Pizza, Nice People!" 07/17/2018



Olympic restaurant

425 reviews

#6 of 101 Results

Examples: CQ vs Dataset

Given the example CQ , and the example collected Dataset D , we have:

$$\text{Etype Coverage } Cov(CQ_c) = \frac{|CQ_c \cap D_c| (2)}{|CQ_c| (4)} = 0.5$$

$$\text{Etype Sparsity } Spr(CQ_c) = 1 - 2 * \frac{|CQ_c \cap D_c| (2)}{|CQ_c| + |D_c| (4 + 11)} = 0.73$$

$$\text{Property Coverage } Cov(CQ_p) = \frac{|CQ_p \cap D_p| (6)}{|CQ_p| (8)} = 0.75$$

$$\text{Property Sparsity } Spr(CQ_p) = 1 - 2 * \frac{|CQ_p \cap D_p| (6)}{|CQ_p| + |D_p| (8 + 40)} = 0.75$$

Notice that:

- Intersection information (coverage) should more likely belongs to common or core category.
- Different information (sparsity) should be core or contextual information.

Contents

1 Evaluation on Inception phase

2 Evaluation on Informal Modeling phase

Evaluation purpose on Informal modelling phase

In **schema level**, we have the proposed informal ER model and a set of CQs. We aim to measure:

- If the proposed informal ER model cover CQs, using metric **coverage**.
- If the proposed informal ER model properly extend CQs, using metric **extensiveness**.

Examples: ER model vs CQs

Given a set of Competency Query (CQ), the coverage (Cov) of the ER model (ER) is:

Etype Coverage
$$Cov(CQ_c) = \frac{|CQ_c \cap ER_c|}{|CQ_c|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

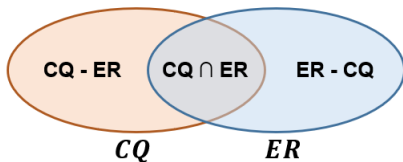
$Cov = 0$ No coverage

Property Coverage
$$Cov(CQ_p) = \frac{|CQ_p \cap ER_p|}{|CQ_p|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

$Cov = 0$ No coverage



Examples: ER model vs CQs

Given a set of Competency Query (CQ) , the Extensiveness (Ext) of the ER model (ER) is:

$$\text{Etype Extensiveness} \quad Ext(CQ_c) = \frac{|ER_c - CQ_c|}{|CQ_c| + |ER_c|}$$

$Ext = 1$ Full Extensiveness

$Ext \simeq 0.5$ Ideal

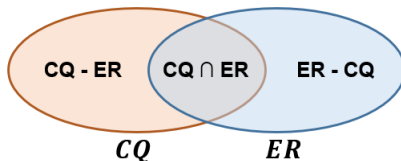
$Ext = 0$ No Extensiveness

$$\text{Property Extensiveness} \quad Ext(CQ_p) = \frac{|ER_p - CQ_p|}{|CQ_p| + |ER_p|}$$

$Ext = 1$ Full Extensiveness

$Ext \simeq 0.5$ Ideal

$Ext = 0$ No Extensiveness



Examples: ER model vs CQs

Given the example CQ, and the example ER model ER, we have:

$$\begin{array}{ll} \text{Etype Coverage} & Cov(CQ_c) = \frac{|CQ_c \cap ER_c| (3)}{|CQ_c| (4)} = 0.75 \\ \text{Etype Extensiveness} & Ext(CQ_c) = \frac{|ER_c - CQ_c| (15)}{|CQ_c| + |ER_c| (4 + 18)} = 0.68 \\ \text{Property Coverage} & Cov(CQ_p) = \frac{|CQ_p \cap ER_p| (5)}{|CQ_p| (8)} = 0.625 \\ \text{Property Extensiveness} & Ext(CQ_p) = \frac{|ER_p - CQ_p| (31)}{|CQ_p| + |ER_p| (8 + 36)} = 0.70 \end{array}$$

Notice that:

- Intersection information (coverage) should more likely belongs to common or core category.
- Additional information (extensiveness) should be core and contextual information. We should find the balance on extensiveness, since too much hard to maintain, too less not properly extend

Evaluation purpose on Informal modelling phase

In **data level**, we have the proposed informal ER model and several collected datasets.

We aim to measure:

- If the informal ER model align with collect datasets, using metric **coverage**.
- If the informal ER model is much different from collect datasets, using metric **Sparsity**.

Examples: ER model vs Dataset

Given the dataset (D), the coverage (Cov) of the ER model (ER) is:

Etype
Coverage

$$Cov(D_c) = \frac{|ER_c \cap D_c|}{|D_c|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

$Cov = 0$ No coverage

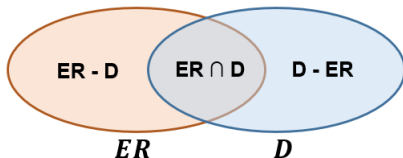
Property
Coverage

$$Cov(D_p) = \frac{|ER_p \cap D_p|}{|D_p|}$$

$Cov = 1$ Full coverage

$0.6 < Cov < 0.8$ Ideal

$Cov = 0$ No coverage



Examples: ER model vs Dataset

Given the dataset (D), the sparsity (Spr) of the ER model (ER) is:

$$\text{Etype Sparsity } Spr(D_c) = \frac{|ER_c - D_c| + |D_c - ER_c|}{|ER_c| + |D_c|} = 1 - 2 * \frac{|ER_c \cap D_c|}{|ER_c| + |D_c|}$$

$Spr = 1$ Full Sparsity

$Spr \approx 0.5$ Ideal

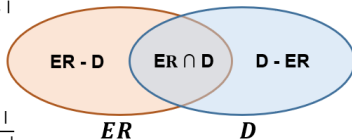
$Spr = 0$ No Sparsity

$$\text{Property Sparsity } Spr(D_p) = \frac{|ER_p - D_p| + |D_p - ER_p|}{|ER_p| + |D_p|} = 1 - 2 * \frac{|ER_p \cap D_p|}{|ER_p| + |D_p|}$$

$Spr = 1$ Full Sparsity

$Spr \approx 0.5$ Ideal

$Spr = 0$ No Sparsity



Examples: ER model vs Dataset

Given the dataset (D), and the example ER model ER , we have:

$$\begin{array}{ll} \text{Etype Coverage} & Cov(D_c) = \frac{|ER_c \cap D_c| (8)}{|D_c| (11)} = 0.73 \\ \text{Etype Sparsity} & Spr(D_c) = 1 - 2 * \frac{|ER_c \cap D_c| (8)}{|ER_c| + |D_c| (16 + 11)} = 0.41 \\ \\ \text{Property Coverage} & Cov(D_p) = \frac{|ER_p \cap D_p| (16)}{|D_p| (40)} = 0.4 \\ \text{Property Sparsity} & Spr(D_p) = 1 - 2 * \frac{|ER_p \cap D_p| (16)}{|ER_p| + |D_p| (36 + 40)} = 0.58 \end{array}$$

Notice that:

- Intersection information (coverage) should more likely belongs to common or core category.
- Different information (sparsity) should be core or contextual information. The sparsity should also keep a balance because if ETG model and dataset are very different, they will be hard to align.



Fausto Giunchiglia



Evaluation

iTelos Inception & Informal
Modeling Phase