



L'ingénierie des documents d'affaires dans le cadre du web sémantique

Jamel Eddine Jridi

Soutenance de thèse, 24 novembre 2014

Introduction



- Échange électronique de documents (depuis 1960).
- Normalisation des échanges.
 - Structuration de l'information.
 - XML comme format d'échange.
- Quelques standards d'affaire basés sur XML:



- **xCBL** (XML Common Business Library).



- **RosettaNet**.



- **cXML** (commerce eXtensible Markup Language).



- **ebXML** (Electronic Business using eXtensible Markup Language).

- ...

Motivation



XML

XML Schema et DTD

Structure arborescente

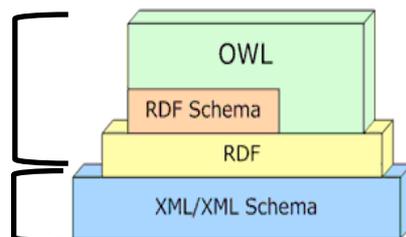
- Imbrication des balises
- Ordre d'apparition

Représentation syntaxique

Validation syntaxique

Sémantique

Syntaxique

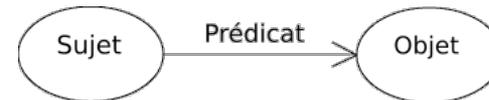


Web sémantique

Web Ontology Language (OWL)

Représentation sous forme de graphe

- Relation entre les concepts



- Hiérarchie des concepts

Représentation sémantique

Vérification de consistance

Capacité d'inférence et de raisonnement
déductif

[Hitzler et al., 2009]



Notre objectif

- Intégrer la sémantique dans les standards d'affaires.
- Représenter les relations entre les concepts.
- Réduire l'hétérogénéité dans les échanges.
- Approches existantes :

- eCl@ssOWL Ontology



- GoodRelations (OWL)



- RosettaNet Ontology (OWL)



- ebXML (OWL-S) [Dogac et al., 2005] et [Schulte et al., 2010]

- RosettaNet (WSML) [Haller et al., 2008] et [Kotinurmi et al., 2009]



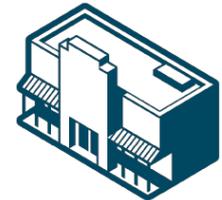
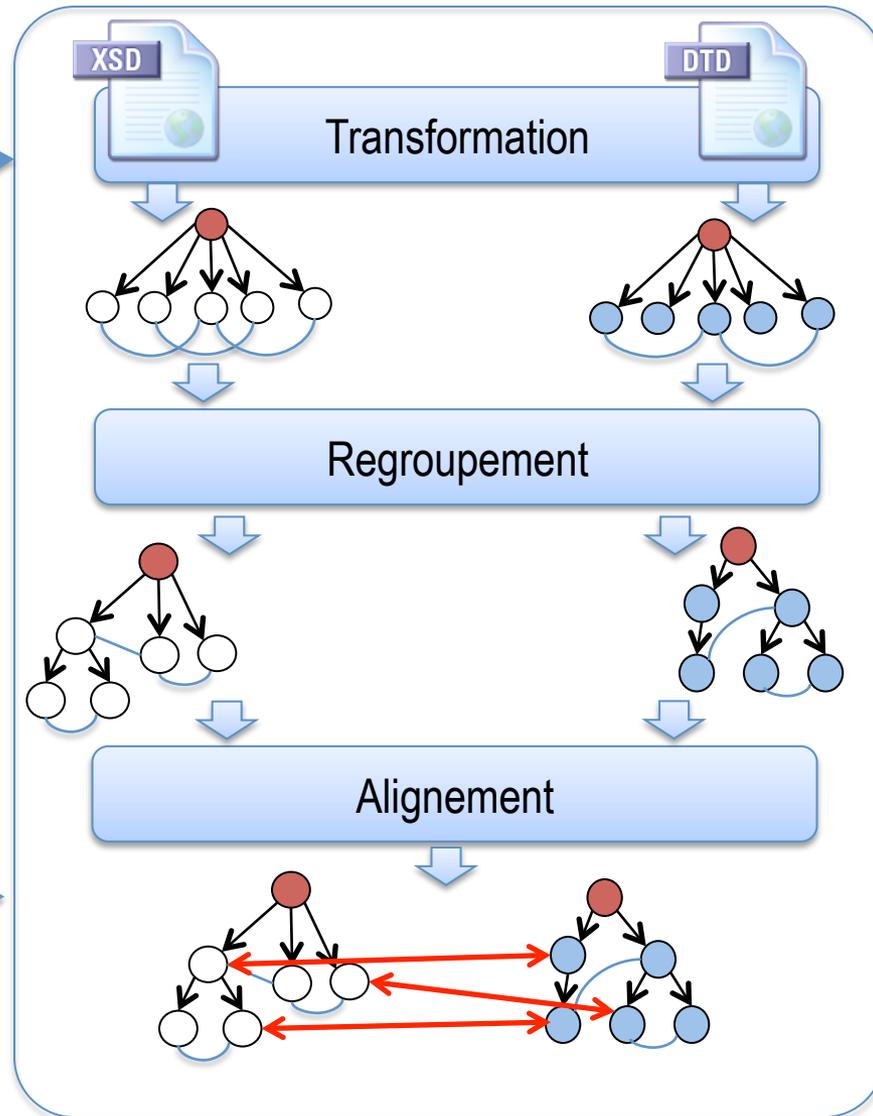
Notre objectif

- Limites:
 - Ne pas profiter de la modélisation XML existante.
 - Nouvelle terminologie.
 - Difficile de convaincre les entreprises.
 - WSMML et OWL-S (Pas de recommandation W3C).
 - Adaptation difficile par les systèmes d'entreprises.
- Solution de rechange:
 - Automatiser la transformation vers OWL.
 - Garder un lien avec ce qui existe déjà.

Approche globale



ROSETTANET
POWERED BY GS1 US

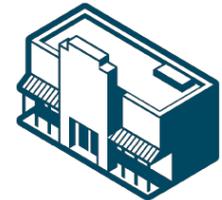
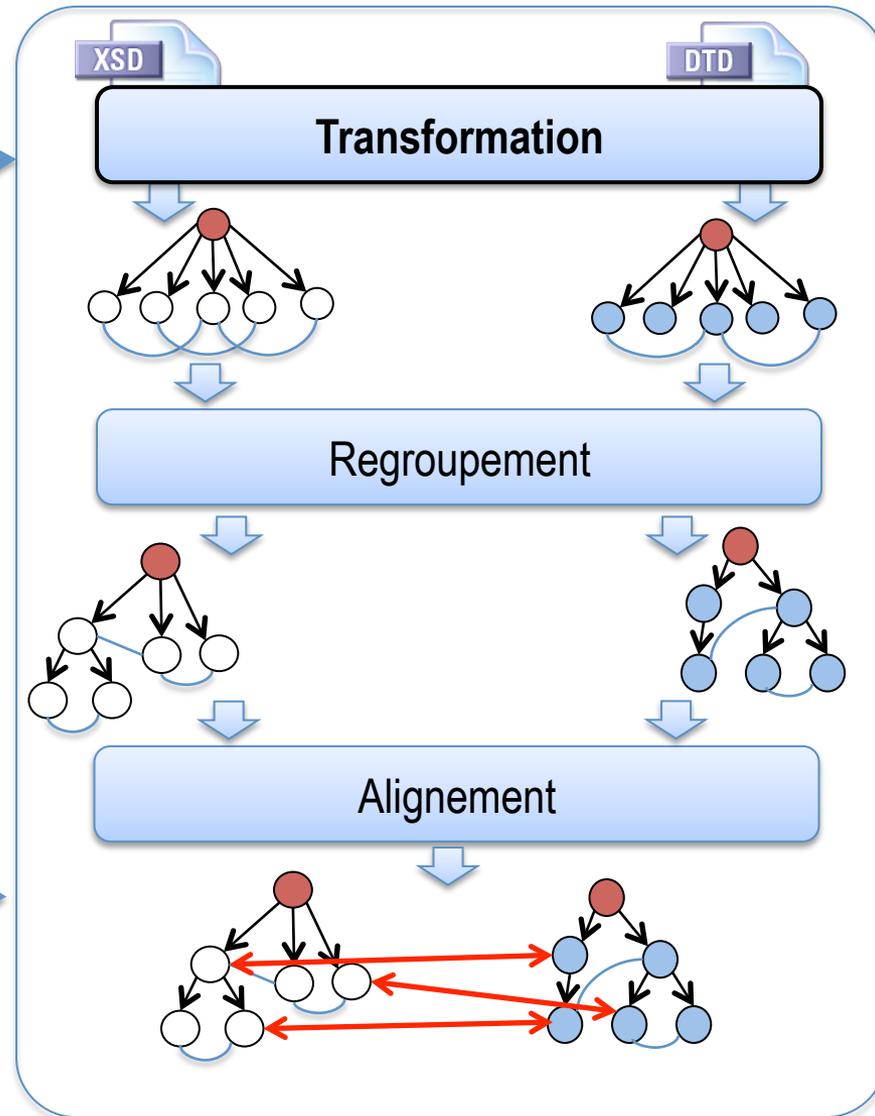


cxm1.org

Approche globale



ROSETTANET
POWERED BY GS1 US



cxm1.org



Étude de l'existant

- ReDeFer [García et al., 2007]
 - XML Schema → RDF.
 - ebXML et BizOntos.
- XS2OWL [Tsinaraki et al., 2007]
 - XML Schema → OWL.
 - Standards Multimédia MPEG-7 et MPEG-21.
- JXML2OWL [Rodrigues et al., 2008]
 - DTD et XML Schema → OWL.

Étude de l'existant



- Certains aspects ne sont pas traités.
 - Les énumérations. (26%)
 - Les annotations. (40%)
 - Les “import”. (2%)
 - Les “namespace”. (5%)
 - Prise en compte des éléments définis dans des fichiers externes.
- Résultat n'était pas valide sémantiquement.
 - Évaluation de la consistance (Pellet).
- JXML2OWL
 - Transformation manuelle.

73%



XSD2OWL

XML Schema → OWL2

[Jridi et al., ICEIS2013]

XML Schema

OWL 2

| | | |
|------------|--|--|
| R1 | complexType group attributeGroup element//complexType | owl:Class |
| R2 | complexType//element | owl:DataProperty owl:ObjectProperty |
| R3 | element/complexType/simpleContent | owl:DataProperty |
| R4 | element@type | owl:DataPropertyRange owl:ObjectPropertyRange |
| R5 | complexType//attribute | owl:DataProperty |
| R6 | attribute@type | owl:DataPropertyRange |
| R7 | enumeration | owl:DataOneOf |
| R8 | @minOccurs | owl:AnnotationAssertion (owl:minCardinality) |
| R9 | @maxOccurs | owl:AnnotationAssertion (owl:maxCardinality) |
| R10 | annotation | owl:AnnotationAssertion (rdfs:comment) |
| R11 | import | owl:Import |
| R12 | namespace | owl:Prefix |



Processus de transformation

```
<xs:complexType name="ProductLineItemType">  
  <xs:sequence>  
    ....  
    <xs:element name="shipTo" type="upi:SpecifiedPartnerDescriptionType"/>  
    ....  
    <xs:element name="Comments" type="xs:string"/>  
    ....  
  </xs:sequence>  
</xs:complexType>
```

XML Schema



OWL 2

```
<owl:Declaration>  
  <owl:Class IRI="ProductLineItemType" />  
</owl:Declaration>
```

```
<owl:Declaration>  
  <owl:ObjectProperty IRI="shipTo" />  
</owl:Declaration>  
  
<owl:ObjectPropertyRange>  
  <owl:ObjectProperty IRI="shipTo" />  
  <owl:Class  
    IRI="&upi;SpecifiedPartnerDescriptionType" />  
</owl:ObjectPropertyRange>  
  
<owl:ObjectPropertyDomain>  
  <owl:ObjectProperty IRI="shipTo" />  
  <owl:Class IRI="ProductLineItemType" />  
</owl:ObjectPropertyDomain>
```

```
<owl:Declaration>  
  <owl:DataProperty IRI="has_Comments" />  
</owl:Declaration>  
  
<owl:DataPropertyRange>  
  <owl:DataProperty IRI="has_Comments" />  
  <owl:Datatype IRI="xs:string" />  
</owl:DataPropertyRange>  
  
<owl:DataPropertyDomain>  
  <owl:DataProperty IRI="has_Comments" />  
  <owl:Class IRI="ProductLineItemType" />  
</owl:DataPropertyDomain>
```



Expérimentation

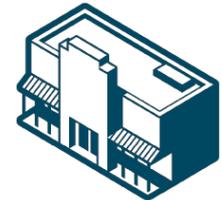
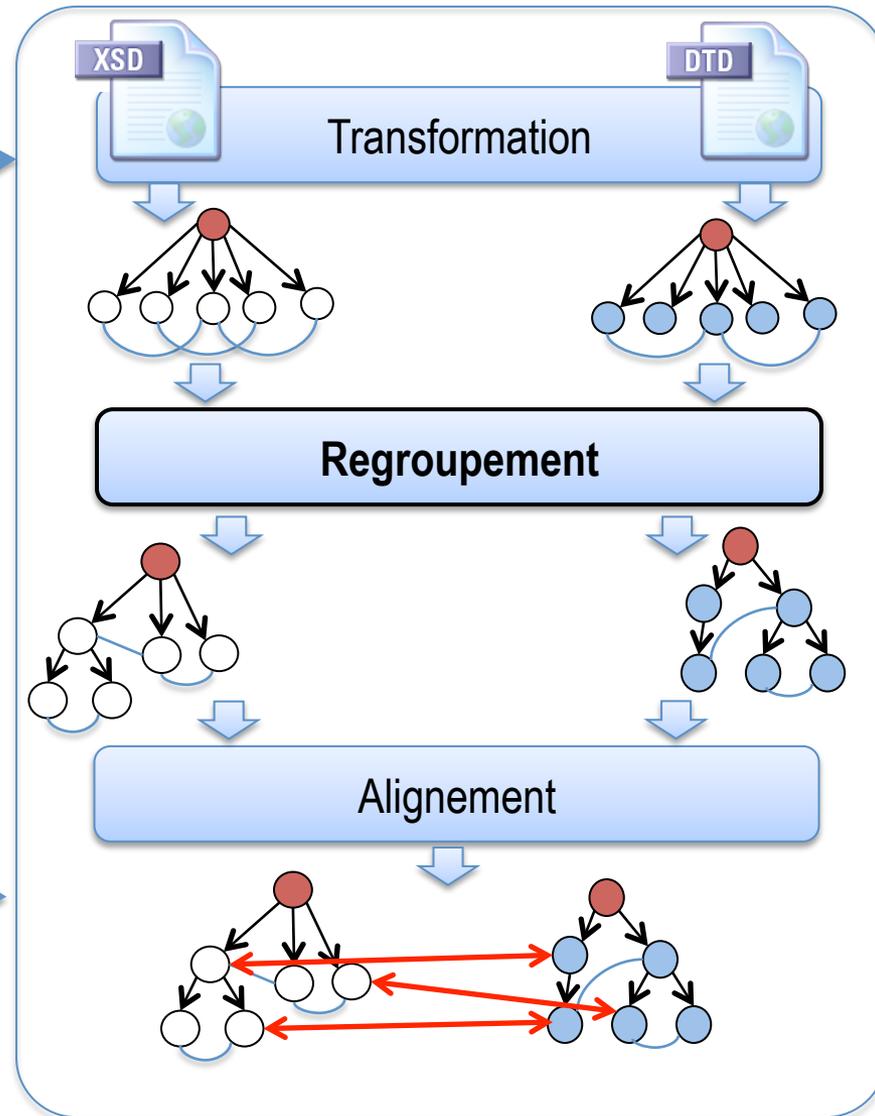
| | RosettaNet | Définition de la métrique |
|-------------|------------|---|
| noc | 1252 | Nombre de classes. |
| nodp | 3045 | Nombre de Data Property. |
| noop | 2607 | Nombre d'Object Property. |
| nop | 5652 | Somme de nodp et noop . |
| nosc | 0 | Nombre de sous-classes. |
| norc | 1252 | Nombre de classes sans superclasses (root classes). |
| nolc | 1252 | Nombre de classes (sans sous-classes). |
| rr | 1 | $\text{nop}/(\text{nop}+\text{nosc})$ |
| ir | 0 | Moyenne du nombre de sous-classes par classe. |

[Sicilia et al., 2012]

Approche globale



ROSETTANET
POWERED BY GS1 US



cxm1.org



Regroupement

[Jridi et al., FCA4AI@IJCAI 2013]

□ But:

- ① Regrouper les classes partageant des propriétés.
- ② Améliorer la représentation de l'ontologie.
- ③ Maintenance plus efficace.

□ Analyse formelle de concep ▼

- Méthode mathématique d'
- Description des relations ϵ
- Basée sur les treillis.

● Thing

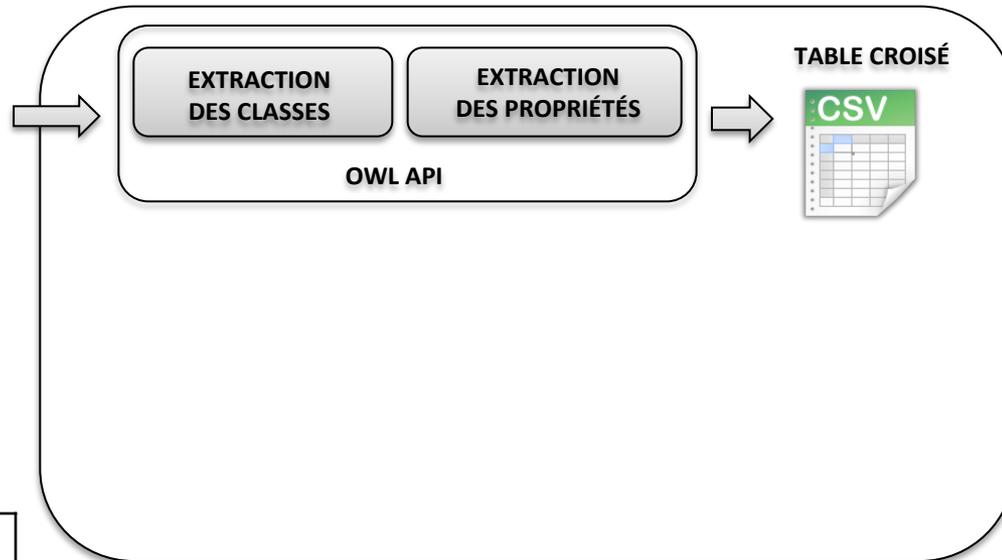
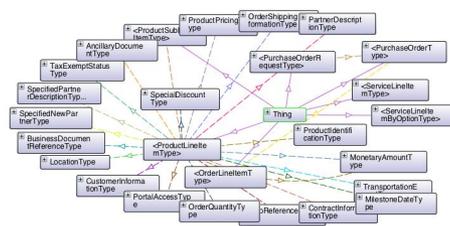
- <OrderLineItemType>
- <ProductLineItemType>
- <ProductSubLineItemType>
- <PurchaseOrderRequestType>
- <PurchaseOrderType>
- <ServiceLineItemByOptionType>
- <ServiceLineItemType>



Regroupement

[Jridi et al. 2013, FCA4AI@IJCAI]

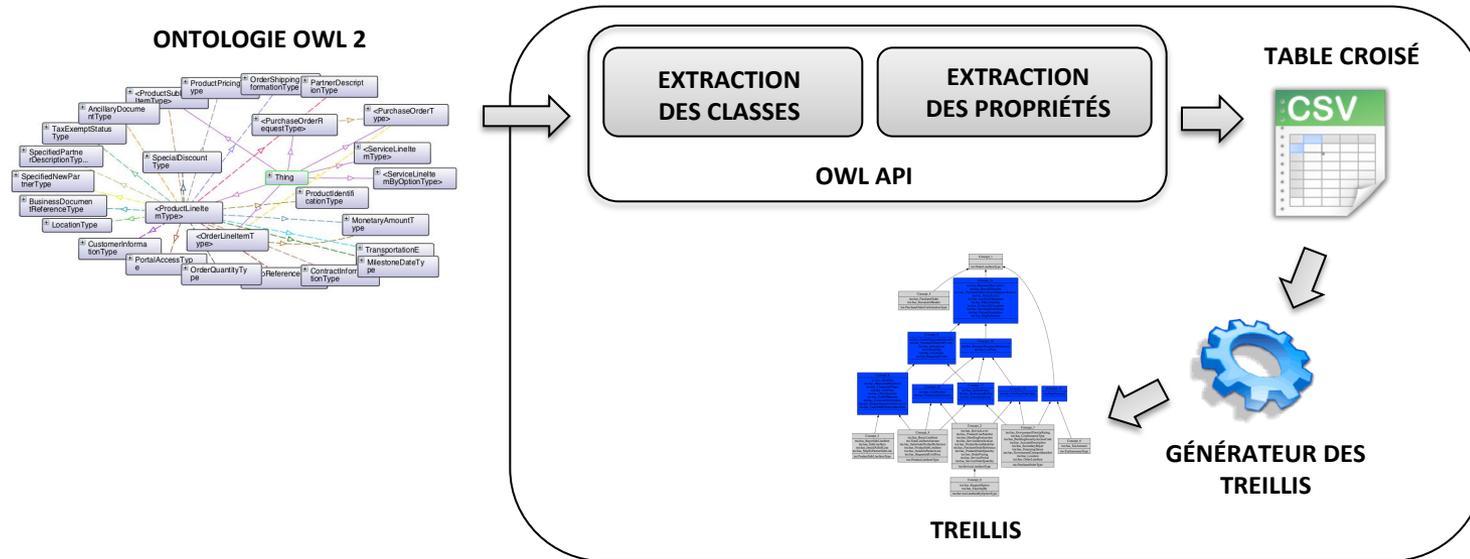
ONTOLOGIE OWL 2



| R | p_1 | p_2 | p_3 | ... | p_m |
|----------|-------|-------|-------|-----|-------|
| O_1 | X | | X | | X |
| O_2 | | X | X | | X |
| ... | | | | | |
| O_k | X | | X | | X |

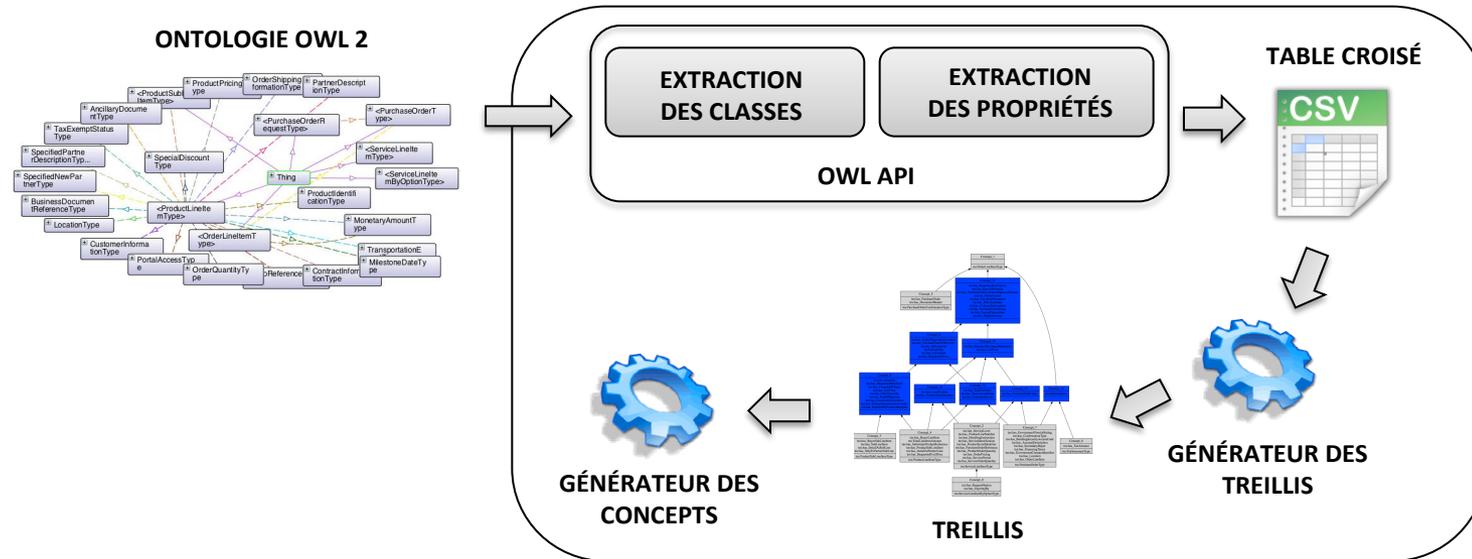
Regroupement

[Jridi et al. 2013, FCA4AI@IJCAI]



Regroupement

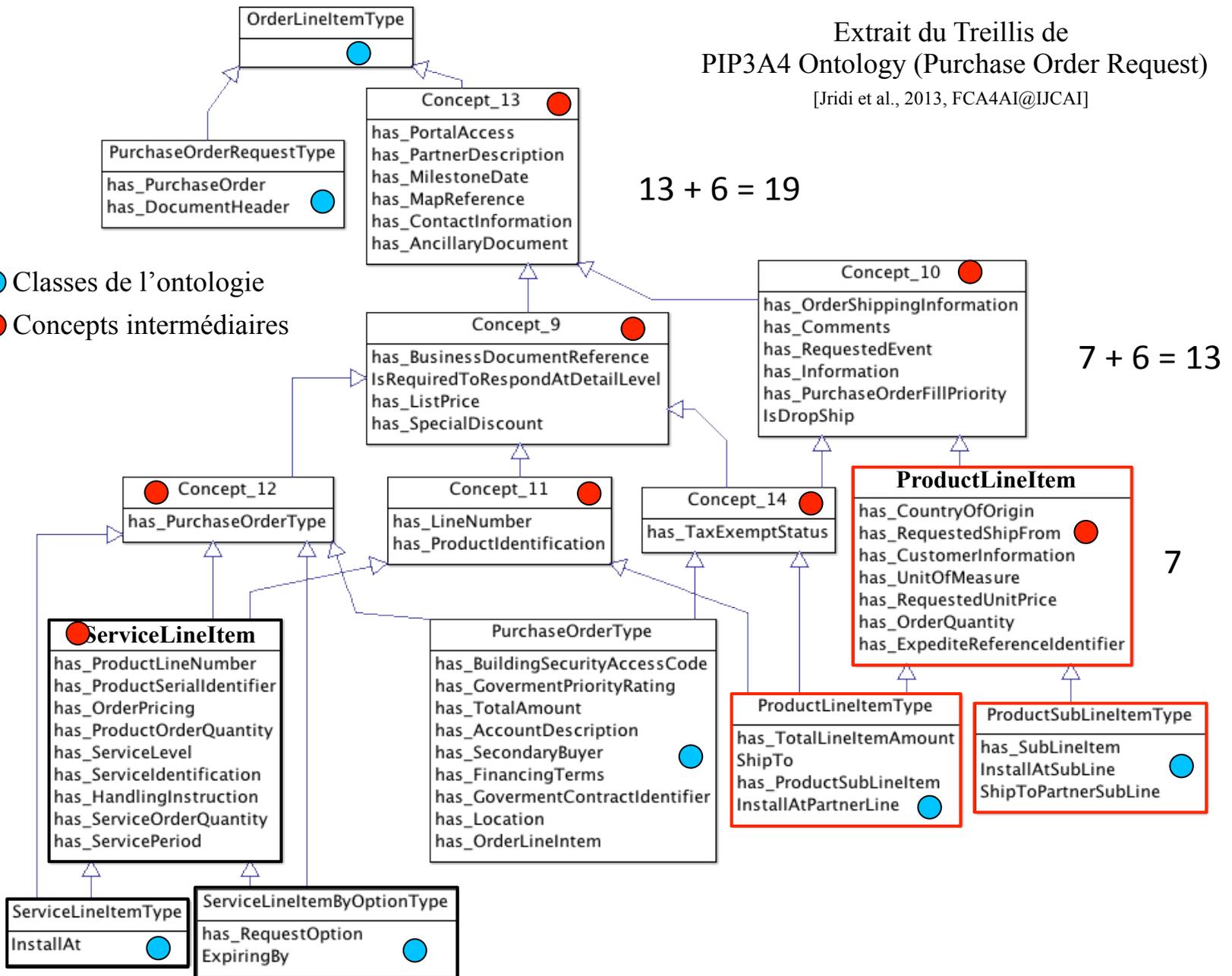
[Jridi et al. 2013, FCA4AI@IJCAI]



Extrait du Treillis de PIP3A4 Ontology (Purchase Order Request)

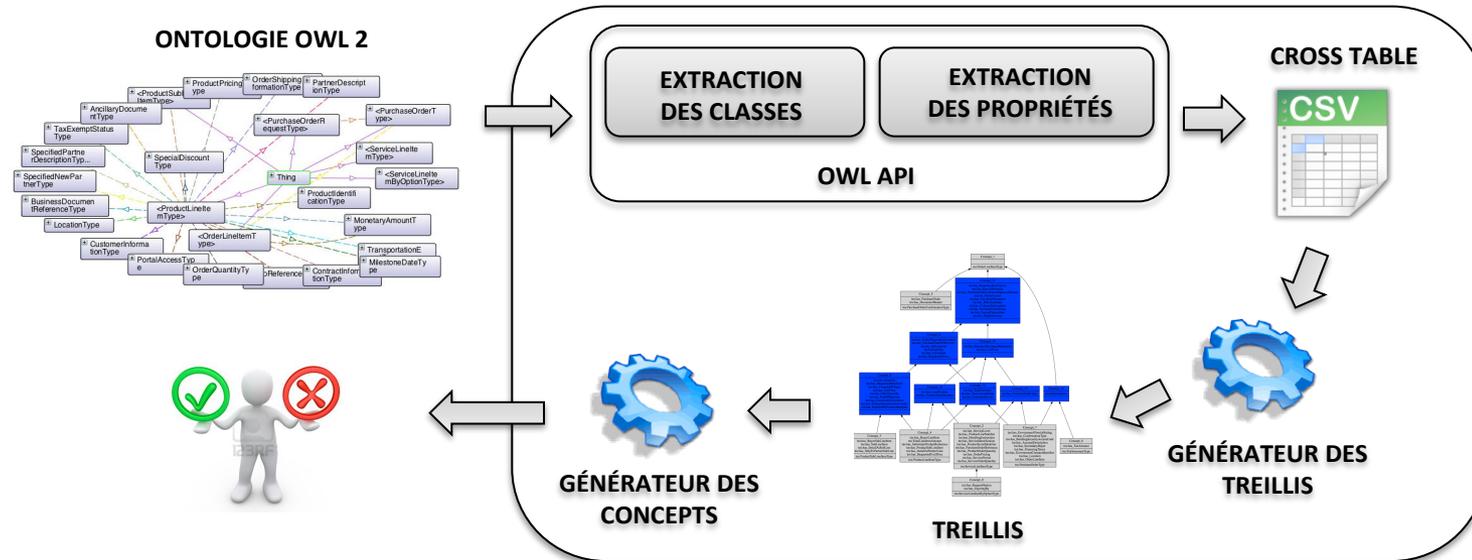
[Jridi et al., 2013, FCA4AI@IJCAI]

- Classes de l'ontologie
- Concepts intermédiaires



Regroupement

[Jridi et al. 2013, FCA4AI@IJCAI]



Expérimentation



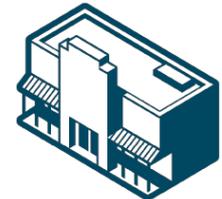
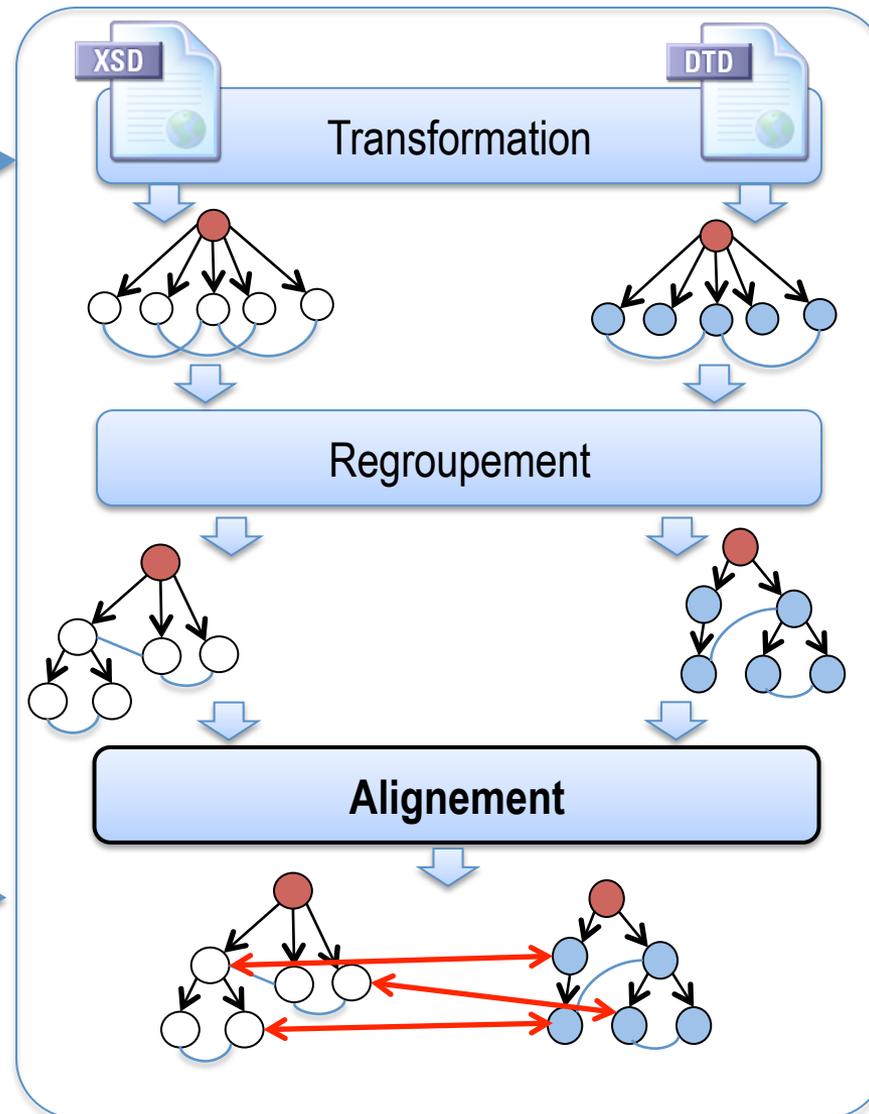
| | Avant | Après | Définition de la métrique |
|-------------|-------|-------------|---|
| noc | 1252 | 1252 | Nombre de classes. |
| nodp | 3045 | 3045 | Nombre de Data Property. |
| noop | 2607 | 2607 | Nombre d'Object Property. |
| nop | 5652 | 5652 | Somme de nodp et noop . |
| nosc | 0 | 384 | Nombre de sous-classes. |
| norc | 1252 | 1050 | Nombre de classes sans superclasses (root classes). |
| nole | 1252 | 868 | Nombre de classes (sans sous-classes). |
| rr | 1 | 0.93 | nop/(nop+nosc) |
| ir | 0 | 0.31 | Moyenne du nombre de sous-classes par classe. |

- Validation manuelle:
 - 90 concepts distincts (concepts partagés entre les documents).
 - 79 ont effectivement une sémantique commune.
 - **Précision : 87%.**

Approche globale

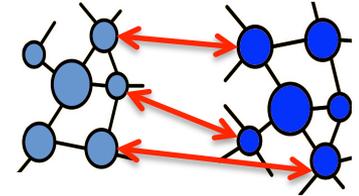


ROSETTANET
POWERED BY GS1 US



cxm1.org

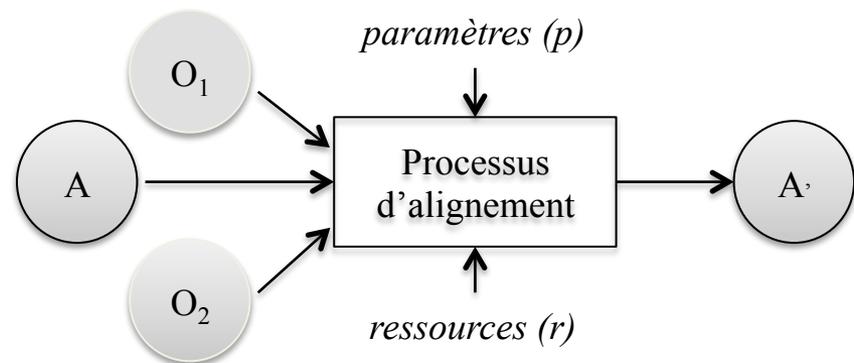
Alignement d'ontologies



« *L'alignement est un processus qui trouve des liens sémantiques ou un ensemble de correspondances entre les entités des ontologies (des classes, des propriétés, etc.).* » [Euzenat et al., 2008]

- Techniques d'alignement :

- Terminologique.
- Structurelle.
- Extensionnelle.
- Sémantique.





Alignement

[Jridi et al., MCETECH2015]

(soumis)

- Techniques terminologiques.
 - Alignement à base de caractères (distance Levenshtein).
 - Alignement à base de tokens (TFIDF).
- Alignement avec OLA_2 [Djougak et al., 2007].
 - Meilleure performance dans les OAEI.
 - Basé sur la structure.
- Ontologies liées à la gestion des bons de commandes.

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad F = 2 * \frac{P * R}{P + R}$$



Alignement

Gestion des bons de commandes

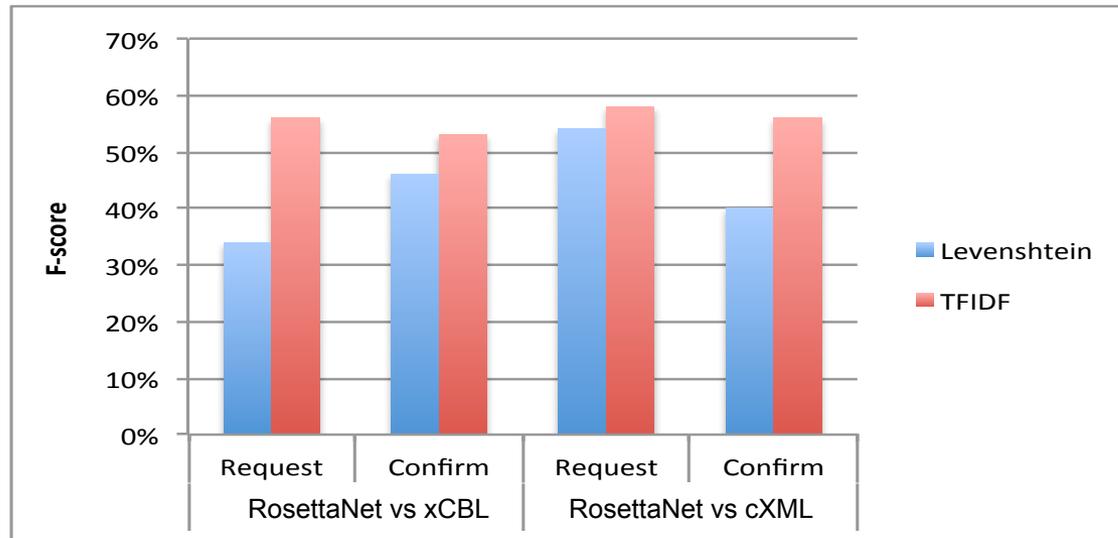
| | Request | Confirm |
|--------------------------|----------------------|---------------------------|
| RosettaNet (.xsd) | PurchaseOrderRequest | PurchaseOrderConfirmation |
| xCBL (.xsd) | OrderRequestType | OrderConfirmationType |
| cXML (.dtd) | OrderRequest | ConfirmationRequest |

| | RosettaNet | | xCBL | | cXML | |
|-------------|-------------------|----------------|----------------|----------------|----------------|----------------|
| | Request | Confirm | Request | Confirm | Request | Confirm |
| noc | 151 | 143 | 66 | 33 | 17 | 27 |
| nodp | 264 | 252 | 174 | 68 | 34 | 36 |
| noop | 126 | 130 | 91 | 35 | 18 | 25 |
| nop | 372 | 382 | 265 | 103 | 52 | 61 |
| nosc | 60 | 58 | 0 | 0 | 0 | 0 |
| norc | 91 | 85 | 66 | 33 | 17 | 27 |
| nolc | 133 | 125 | 66 | 33 | 17 | 27 |
| rr | 0.86 | 0.86 | 1 | 1 | 1 | 1 |
| ir | 0.40 | 0.40 | 0 | 0 | 0 | 0 |



Alignement

Techniques terminologiques



- Importance des annotations
- Les ontologies partagent les termes les plus fréquents.

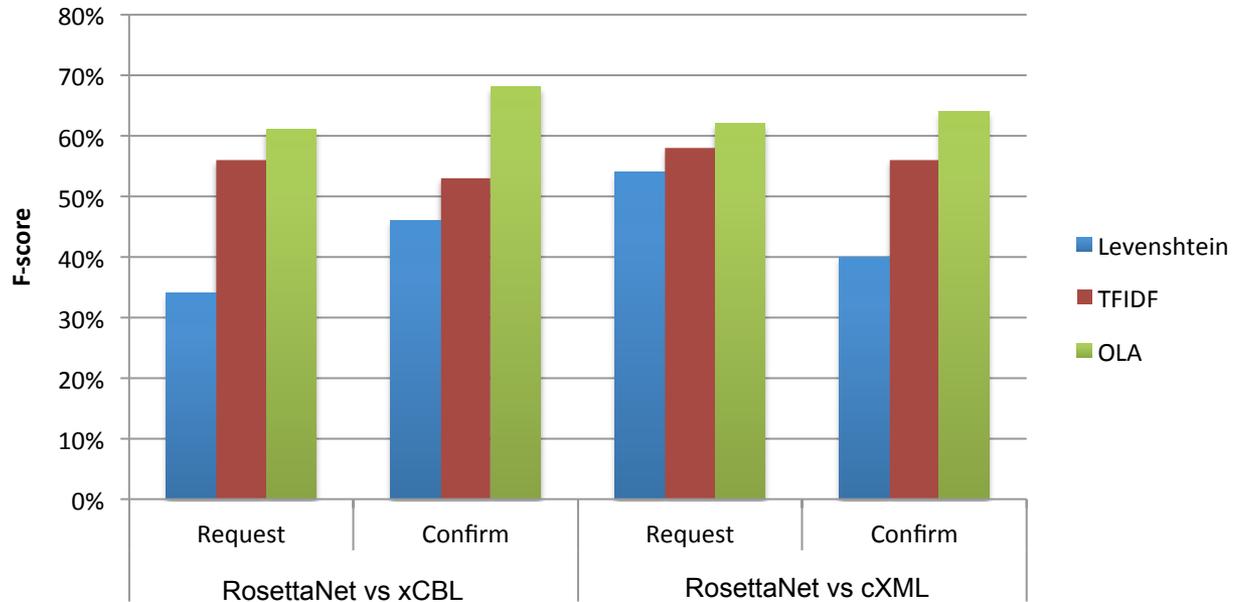
“OrderRequest” de cXML

| | TFIDF | Levenshtein |
|---|-------|-------------|
| RequestingOrderLineItemReferenceType (RosettaNet) | 0.59 | 0.25 |
| → PurchaseOrderRequestType (RosettaNet) | 0.49 | 0.50 |



Alignement

OLA₂ [Djoufak et al., 2007]



Alignement de "StreetSupplement2" (xCBL)

Vrai/Faux

| | | |
|------------------|-----------------------------|------|
| TFIDF | Information (RosettaNet) | Faux |
| OLA ₂ | AddressLine2 (RosettaNet) | Vrai |
| Levenshtein | TransportEvent (RosettaNet) | Faux |





Conclusion

- Intégration de la sémantique dans les standards d'affaire.
- Interopérabilité efficace.
- Transformation des documents en DTD et XML Schema vers OWL 2.
- FCA : regroupement des concepts ayant une sémantique commune.
- Réduire l'hétérogénéité à l'aide d'alignements.



Travaux futurs

- Étudier d'autres relations sémantiques à partir du XML Schema (p.ex. les classes disjointes).
- Essayer d'autres techniques d'alignement :
 - Extensionnelle en présence d'instances de documents.
 - Sémantique en présence d'un thésaurus du domaine.
- Concevoir un thésaurus pour répertorier les différents termes techniques liés au domaine d'affaires.

Merci de votre attention