

Intégration de données environnementales : une approche basée sur les entrepôts de documents XML et les ontologies

Ousmane Sall^{1 2} Moussa Lô²

¹Laboratoire d'Analyse Numérique et d'Informatique - UFR SAT
Université Gaston Berger
Saint-Louis(Sénégal)
lom@ugb.sn

²Laboratoire d'Informatique du Littoral
Université du Littoral Côte d'Opale
Calais(France)
sall@lil.univ-littoral.fr

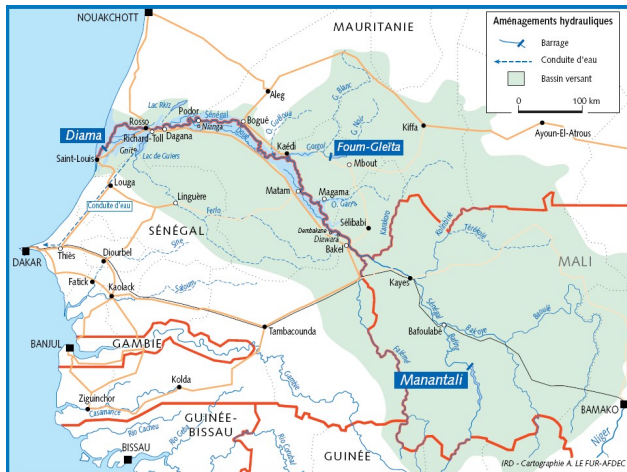
Plan

- 1 Problématique et contexte
 - Contexte
 - Problématique
- 2 Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- 3 Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- 4 Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Contexte



Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Hétérogénéité structurelle et sémantique des données

- Distribuées sur plusieurs sources appelées **partenaires**
- De nature spatio-temporelle, i.e **recueillies dans un espace précis et selon une échelle de temps définie**
- Vocabulaire propre (**hétérogénéité sémantique**) à chaque partenaire
- Des moyens et structures de représentation propres (**hétérogénéité structurelle**) à chaque partenaire

Thématiques diverses selon les partenaires

- SAED, OMS, ADRAO, DRDR/SL,...
- Santé, Agriculture, Elevage, Pêche, Hydrographie...

Problématique

Problème induisant un besoin d'intégration

- Gestion et exploitation des données optimales...
- Coopération entre partenaires, avec contraintes : une exploitation individuelle et autonome des données (**nature confidentielle de certaines données**)

Exemple (Exemple de requête)

Années pour lesquelles la **pluviométrie moyenne** est supérieure ou égale à 30 dans la **région de Saint-Louis** et la **production de riz** supérieure à 5 tonnes à l'hectare avec **invasion d'oiseaux granivores** et un **taux de remboursement des dettes des paysans cultivant le riz** dépassant 50% dans le **département de Saint-Louis** ?

Problématique

Problème induisant un besoin d'intégration

- Gestion et exploitation des données optimales...
- Coopération entre partenaires, avec contraintes : une exploitation individuelle et autonome des données (**nature confidentielle de certaines données**)

Exemple (Exemple de requête)

Années pour lesquelles la **pluviométrie moyenne** est supérieure ou égale à 30 dans la **région de Saint-Louis** et la **production de riz** supérieure à 5 tonnes à l'hectare avec **invasion d'oiseaux granivores** et un **taux de remboursement des dettes des paysans cultivant le riz** dépassant 50% dans le **département de Saint-Louis** ?

Problématique

Problème induisant un besoin d'intégration

- Gestion et exploitation des données optimales...
- Coopération entre partenaires, avec contraintes : une exploitation individuelle et autonome des données (**nature confidentielle de certaines données**)

Exemple (Exemple de requête)

Années pour lesquelles la **pluviométrie moyenne** est supérieure ou égale à 30 dans la **région de Saint-Louis** et la **production de riz** supérieure à 5 tonnes à l'hectare avec **invasion d'oiseaux granivores** et un **taux de remboursement des dettes des paysans cultivant le riz** dépassant 50% dans le **département de Saint-Louis** ?

Plan

- 1 Problématique et contexte
 - Contexte
 - Problématique
- 2 Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- 3 Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- 4 Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Construction automatique : trois classes de méthodes

Généralement celle d'Uschold et Gruninger (1996) usée.

Techniques statistiques et d'apprentissage automatique

Kietzet et al.(2000),Maedche et Staab(2001),Navigli et al.(2003)

Techniques basées sur la migration d'un thésaurus en ontologie

Miles et al.(2003), Wielinga et al.(2001) ou Clark et al.(2000) :
réutilisation d'une ontologie ou une hiérarchie de concepts
comme base de connaissance

Les méthodes basées sur des textes non structurés

TERMINAE(Aussenac-Gilles et al. (2000), Biébow et al.
(2000)) reposant sur l'analyse de corpus linguistiques

Construction automatique : trois classes de méthodes

Généralement celle d'Uschold et Gruninger (1996) usée.

Techniques statistiques et d'apprentissage automatique

Kietzet et al.(2000),Maedche et Staab(2001),Navigli et al.(2003)

Techniques basées sur la migration d'un thésaurus en ontologie

Miles et al.(2003), Wielinga et al.(2001) ou Clark et al.(2000) :
réutilisation d'une ontologie ou une hiérarchie de concepts
comme base de connaissance

Les méthodes basées sur des textes non structurés

TERMINAE(Aussenac-Gilles et al. (2000), Biébow et al.
(2000)) reposant sur l'analyse de corpus linguistiques

Construction automatique : trois classes de méthodes

Généralement celle d'Uschold et Gruninger (1996) usée.

Techniques statistiques et d'apprentissage automatique

Kietzet et al.(2000),Maedche et Staab(2001),Navigli et al.(2003)

Techniques basées sur la migration d'un thésaurus en ontologie

Miles et al.(2003), Wielinga et al.(2001) ou Clark et al.(2000) :
réutilisation d'une ontologie ou une hiérarchie de concepts
comme base de connaissance

Les méthodes basées sur des textes non structurés

TERMINAE(Aussenac-Gilles et al. (2000), Biébow et al.
(2000)) reposant sur l'analyse de corpus linguistiques

Construction automatique : trois classes de méthodes

Généralement celle d'Uschold et Gruninger (1996) usée.

Techniques statistiques et d'apprentissage automatique

Kietzet et al.(2000),Maedche et Staab(2001),Navigli et al.(2003)

Techniques basées sur la migration d'un thésaurus en ontologie

Miles et al.(2003), Wielinga et al.(2001) ou Clark et al.(2000) :
réutilisation d'une ontologie ou une hiérarchie de concepts
comme base de connaissance

Les méthodes basées sur des textes non structurés

TERMINAE(Aussenac-Gilles et al. (2000), Biébow et al.
(2000)) reposant sur l'analyse de corpus linguistiques

Construction automatique : trois classes de méthodes

Généralement celle d'Uschold et Gruninger (1996) usée.

Techniques statistiques et d'apprentissage automatique

Kietzet et al.(2000),Maedche et Staab(2001),Navigli et al.(2003)

Techniques basées sur la migration d'un thésaurus en ontologie

Miles et al.(2003), Wielinga et al.(2001) ou Clark et al.(2000) :
réutilisation d'une ontologie ou une hiérarchie de concepts
comme base de connaissance

Les méthodes basées sur des textes non structurés

TERMINAE(Aussenac-Gilles et al. (2000), Biébow et al.
(2000)) reposant sur l'analyse de corpus linguistiques

Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Ontologie basée sur la structure des documents XML

Intégration de documents XML

- Klein(2002) : transformation directe des données XML en RDF
- Patel-Schneider et Siméon (Clark et al.2000) : **interrogation de sources hétérogènes, avec des syntaxes et modèles différents**
- Lakshmanan et Sadri(2003) : **fondée sur la disponibilité d'une ontologie standard spécifique à l'application servant de schéma global**
- Une méthodologie de mapping de XML à RDF et aussi de XML-Schema à OWL est proposée dans Matthias (2004)
- Reif et al.(2005) et Steve(2004) : WEESA. **Les règles de mapping de XML-Schema à une ontologie OWL soient générées manuellement**
- Hannes(2004) : relations structurelles uniquement
- Isabel(2004) : Génération automatique d'une ontologie OWL pour chaque document et fusion pour la construction d'une ontologie globale à la source

Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Architecture et construction acedente



Fig.: Architecture du SIC-Sénégal

Deux approches :

- 1 Entrepôt : Intégration des données partenaires
- 2 Médiateur : Intégration des partenaires

Intégration en trois phases :

- 1 Pré-intégration
- 2 Construction des ontologies locales OWL
- 3 Construction de l'ontologie globale OWL

Architecture et construction acedente

Deux approches :

- ① **Entrepôt : Intégration des données partenaires**
- ② **Médiateur : Intégration des partenaires**

Intégration en trois phases :

- ① **Pré-intégration**
- ② **Construction des ontologies locales OWL**
- ③ **Construction de l'ontologies globale OWL**

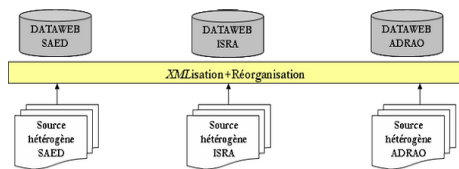


Fig.: Architecture du SIC-Sénégal

Architecture et construction acedente

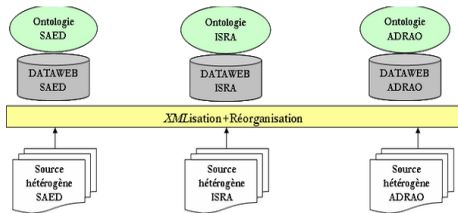


Fig.: Architecture du SIC-Sénégal

Deux approches :

- ① Entrepôt : Intégration des données partenaires
- ② Médiateur : Intégration des partenaires

Intégration en trois phases :

- ① Pré-intégration
- ② Construction des ontologies locales OWL
- ③ Construction de l'ontologies globale OWL

Architecture et construction acedente

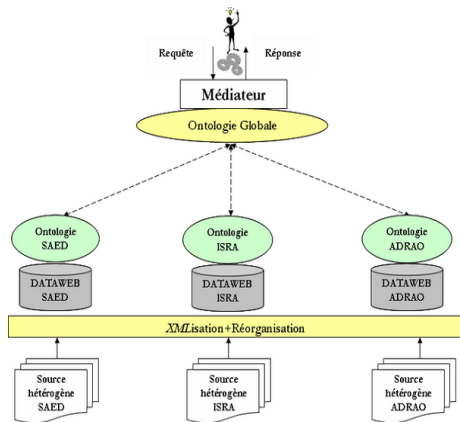


Fig.: Architecture du SIC-Sénégal

Deux approches :

- 1 Entrepôt : Intégration des données partenaires
- 2 **Médiateur : Intégration des partenaires**

Intégration en trois phases :

- 1 Pré-intégration
- 2 Construction des ontologies locales OWL
- 3 **Construction de l'ontologies globale OWL**

Architecture et construction acedente

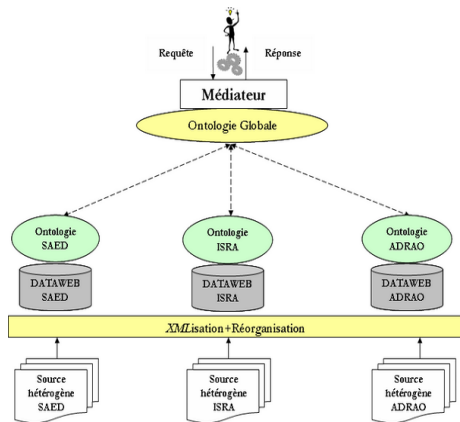


Fig.: Architecture du SIC-Sénégal

Deux approches :

- ① Entrepôt : Intégration des données partenaires
- ② **Médiateur : Intégration des partenaires**

Intégration en trois phases :

- ① Pré-intégration
- ② Construction des ontologies locales OWL
- ③ **Construction de l'ontologies globale OWL**

Construction des dataweb partenaires

Statistiques d'exportation		
Tomates séchées et haricots verts		
Années	Tonnage exporté	
	Tomates séchées	Haricots verts extra fins
1997	100	150 à 200
1998	100	
1999	100	400
2000	100	400
2001	100	400
2002	100	400

```
<Riz>
  <_Sup____ha_> 28 371</_Sup____ha_>
  <_Prod__T__> 144 875 </_Prod__T__>
  <_Rendt__T__ha_> 5 </_Rendt__T__ha_>
</Riz>
```

Après réorganisation:

```
<riz>
  <superficie valeur="28 371" unite_de_mesure="ha" />
  <production valeur="144 875" unite_de_mesure="t" />
  <rendement valeur="5" unite_de_mesure="t_ha" />
</riz>
```

Constitution des entrepôts de documents XML ou Dataweb

- 1 Intégration structurelle (migration des données vers XML)
- 2 Intégration spatio-temporelle (réorganisation des documents et mise en exergue des caractéristiques spatio-temporelles)
- 3 Utilisation d'un wrapper Excel-XML (Kasset et Niang (2006))

Construction des dataweb partenaires

Statistiques d'exportation		
Tomates séchées et haricots verts		
Années	Tonnage exporté	
	Tomates séchées	Haricots verts extra fins
1997	100	150 à 200
1998	100	
1999	100	400
2000	100	400
2001	100	400
2002	100	400

```
<Riz>
  <Sup____ha_> 28 371</ Sup____ha_>
  <Prod_T_> 144 875 </Prod_T_>
  <_Rendt_T_ha_> 5 </_Rendt_T_ha_>
</Riz>
```

Après réorganisation:

```
<riz>
  <superficie valeur="28 371" unite_de_mesure="ha" />
  <production valeur="144 875" unite_de_mesure="t" />
  <rendement valeur="5" unite_de_mesure="t_ha" />
</riz>
```

Constitution des entrepôts de documents XML ou Dataweb

- 1 Intégration structurelle (migration des données vers XML)
- 2 Intégration spatio-temporelle (réorganisation des documents et mise en exergue des caractéristiques spatio-temporelles)
- 3 Utilisation d'un wrapper Excel-XML (Kasset et Niang (2006))

Construction des dataweb partenaires

Statistiques d'exportation		
Tomates séchées et haricots verts		
Années	Tonnage exporté	
	Tomates séchées	Haricots verts extra fins
1997	100	150 à 200
1998	100	
1999	100	400
2000	100	400
2001	100	400
2002	100	400

```
<Riz>
  <Sup____ha_> 28 371</ Sup____ha_>
  <Prod_T_> 144 875 </Prod_T_>
  < Rendt_T_ha_> 5 </ Rendt_T_ha_>
</Riz>
```

Après réorganisation:

```
<riz>
  <superficie valeur="28 371" unite_de_mesure="ha" />
  <production valeur="144 875" unite_de_mesure="t" />
  <rendement valeur="5" unite_de_mesure="t_ha" />
</riz>
```

Constitution des entrepôts de documents XML ou Dataweb

- 1 Intégration structurelle (migration des données vers XML)
- 2 Intégration spatio-temporelle (réorganisation des documents et mise en exergue des caractéristiques spatio-temporelles)
- 3 Utilisation d'un wrapper Excel-XML (Kasset et Niang (2006))

Construction des dataweb partenaires

Statistiques d'exportation		
Tomates séchées et haricots verts		
Années	Tonnage exporté	
	Tomates séchées	Haricots verts extra fins
1997	100	150 à 200
1998	100	
1999	100	400
2000	100	400
2001	100	400
2002	100	400

```
<Riz>
  <Sup____ha_> 28 371</ Sup____ha_>
  <Prod_T_> 144 875 </Prod_T_>
  <_Rendt_T_ha_> 5 </_Rendt_T_ha_>
</Riz>
```

Après réorganisation:

```
<riz>
  <superficie valeur="28 371" unite_de_mesure="ha" />
  <production valeur="144 875" unite_de_mesure="t" />
  <rendement valeur="5" unite_de_mesure="t_ha" />
</riz>
```

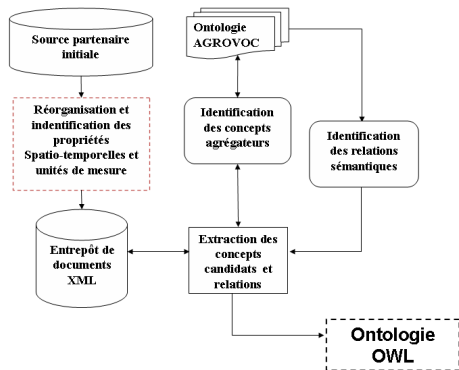
Constitution des entrepôts de documents XML ou Dataweb

- 1 Intégration structurelle (migration des données vers XML)
- 2 Intégration spatio-temporelle (réorganisation des documents et mise en exergue des caractéristiques spatio-temporelles)
- 3 Utilisation d'un wrapper Excel-XML (Kasset et Niang (2006))

Plan

- 1 Problématique et contexte
 - Contexte
 - Problématique
- 2 Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- 3 Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- 4 Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Intégration sémantique des données (chaque partenaire)



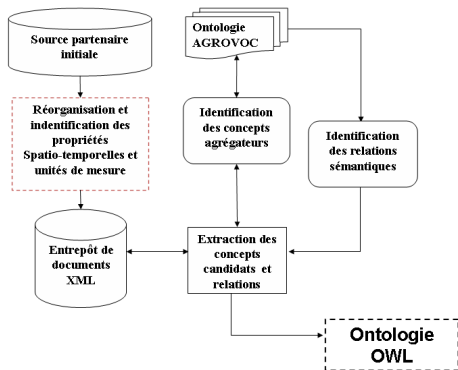
AGROVOC de la FAO

- Ontologie OWL multilingue
- Taxonomie d'un ensemble d'environ 36.000 concepts **environnementaux**
- Plusieurs domaines : pêche, agriculture, élevage, santé, alimentation, géographie,...
- Couvre les domaines du SIC sauf celui pluviométrique

Phase d'intégration sémantique (niveau partenaire)

Extraction et agrégation à AGROVOC de chaque concepts candidats

Intégration sémantique des données (chaque partenaire)



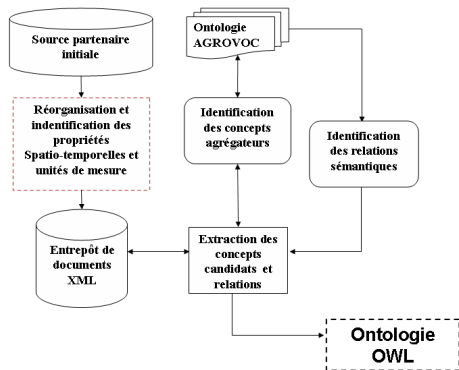
AGROVOC de la FAO

- Ontologie OWL multilingue
- Taxonomie d'un ensemble d'environ 36.000 concepts **environnementaux**
- Plusieurs domaines : pêche, agriculture, élevage, santé, alimentation, géographie,...
- Couvre les domaines du SIC sauf celui pluviométrique

Phase d'intégration sémantique (niveau partenaire)

Extraction et agrégation à AGROVOC de chaque concepts candidats

Intégration sémantique des données (chaque partenaire)



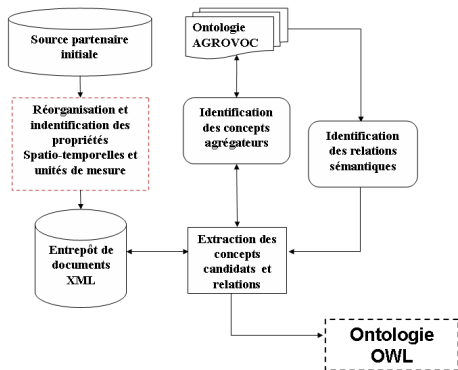
AGROVOC de la FAO

- Ontologie OWL multilingue
- Taxonomie d'un ensemble d'environ 36.000 concepts **environnementaux**
- Plusieurs domaines : pêche, agriculture, élevage, santé, alimentation, géographie,...
- Couvre les domaines du SIC sauf celui pluviométrique

Phase d'intégration sémantique (niveau partenaire)

Extraction et agrégation à AGROVOC de chaque concepts candidats

Intégration sémantique des données (chaque partenaire)



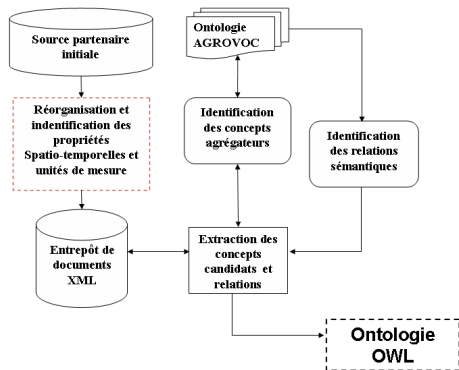
AGROVOC de la FAO

- Ontologie OWL multilingue
- Taxonomie d'un ensemble d'environ 36.000 concepts **environnementaux**
- Plusieurs domaines : pêche, agriculture, élevage, santé, alimentation, géographie,...
- Couvre les domaines du SIC sauf celui pluviométrique

Phase d'intégration sémantique (niveau partenaire)

Extraction et agrégation à AGROVOC de chaque concepts candidats

Intégration sémantique des données (chaque partenaire)



AGROVOC de la FAO

- Ontologie OWL multilingue
- Taxonomie d'un ensemble d'environ 36.000 concepts **environnementaux**
- Plusieurs domaines : pêche, agriculture, élevage, santé, alimentation, géographie,...
- Couvre les domaines du SIC sauf celui pluviométrique

Phase d'intégration sémantique (niveau partenaire)

Extraction et agrégation à AGROVOC de chaque concepts candidats

Extraction des concepts candidats

Extraction des concepts candidats

L'**origine tabloïdale** des données permet d'opter pour le schéma de mapping suivant :

Mapping XML à OWL

Composant XML	Composant OWL
Nœud non terminal	Classe
Nœud terminal avec attributs	Classe
Nœud terminal sans attribut	Propriété de son nœud père
Attribut de nœud	Propriété de son nœud père

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Agrégation, Extraction des relations et Nettoyage

Agrégation des concepts candidats

- ➊ Après extraction de chaque concept candidat : recherche d'un concept agrégateur dans AGROVOC
- ➋ Apport sémantique : grâce au concept subsumé dans AGROVOC

Extraction des relations sémantiques

- ➊ Relations hiérarchiques issues de la structure du document XML
- ➋ Relations sémantiques comme la subsumption : héritée du concept subsumé dans AGROVOC
- ➌ Utilisation de la propriété multi-langue d'Agrovoc pour la recherche de relations supplémentaires.

Nettoyage des différentes ontologies

Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Problématique et objectif de départ

- Dans Contexte du SIC-Sénégal : offrir une plate-forme permettant à des partenaires de partager et d'interroger des sources de données hétérogènes
- Hétérogénéité structurelle, hétérogénéité sémantique avec un vocabulaire commun à chaque partenaire ou partagé

Plusieurs particularités

- Nature environnementale et donc spatio-temporelle
- Une représentation sous des structures différentes selon les partenaires
- Contrainte des partenaires : une exploitation individuelle et autonome de leurs données (Système de négociation)

Conclusion

Approche basée sur les entrepôts

- Entrepôt de documents XML ou dataweb : palier à l'hétérogénéité structurelle et l'aspect propriétaire des données au sein de chaque partenaire
- Intégration des différents entrepôts : association d'une ontologie OWL à chaque dataweb partenaire

Approche basée sur les ontologies

- Les ontologies sont construites automatiquement par une extraction des concepts candidats et leurs termes ou labels à partir de la structure des documents XML
- En plus des relations hiérarchiques, inférence des relations sémantiques telles que la subsumption grâce à AGROVOC

Conclusion

Approche basée sur les entrepôts

- Entrepôt de documents XML ou dataweb : palier à l'hétérogénéité structurelle et l'aspect propriétaire des données au sein de chaque partenaire
- Intégration des différents entrepôts : association d'une ontologie OWL à chaque dataweb partenaire

Approche basée sur les ontologies

- Les ontologies sont construites automatiquement par une extraction des concepts candidats et leurs termes ou labels à partir de la structure des documents XML
- En plus des relations hiérarchiques, inférence des relations sémantiques telles que la subsumption grâce à AGROVOC

Conclusion

Approche basée sur les entrepôts

- Entrepôt de documents XML ou dataweb : palier à l'hétérogénéité structurelle et l'aspect propriétaire des données au sein de chaque partenaire
- Intégration des différents entrepôts : association d'une ontologie OWL à chaque dataweb partenaire

Approche basée sur les ontologies

- Les ontologies sont construites automatiquement par une extraction des concepts candidats et leurs termes ou labels à partir de la structure des documents XML
- En plus des relations hiérarchiques, inférence des relations sémantiques telles que la subsumption grâce à AGROVOC

Conclusion

Approche basée sur les entrepôts

- Entrepôt de documents XML ou dataweb : palier à l'hétérogénéité structurelle et l'aspect propriétaire des données au sein de chaque partenaire
- Intégration des différents entrepôts : association d'une ontologie OWL à chaque dataweb partenaire

Approche basée sur les ontologies

- Les ontologies sont construites automatiquement par une extraction des concepts candidats et leurs termes ou labels à partir de la structure des documents XML
- En plus des relations hiérarchiques, inférence des relations sémantiques telles que la subsumption grâce à AGROVOC

Conclusion

Approche basée sur les entrepôts

- Entrepôt de documents XML ou dataweb : palier à l'hétérogénéité structurelle et l'aspect propriétaire des données au sein de chaque partenaire
- Intégration des différents entrepôts : association d'une ontologie OWL à chaque dataweb partenaire

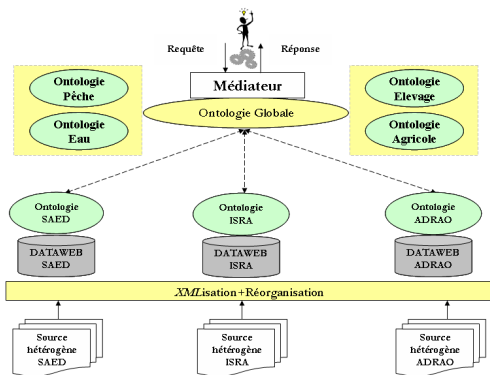
Approche basée sur les ontologies

- Les ontologies sont construites automatiquement par une extraction des concepts candidats et leurs termes ou labels à partir de la structure des documents XML
- En plus des relations hiérarchiques, inférence des relations sémantiques telles que la subsumption grâce à AGROVOC

Plan

- ① Problématique et contexte
 - Contexte
 - Problématique
- ② Existant en construction automatique d'ontologie
 - Construction automatique d'ontologie
 - Construction d'ontologie à partir de documents XML
- ③ Une approche basée sur les entrepôts de documents XML et les ontologies
 - Une approche entrepôt XML
 - Construction automatique d'ontologie OWL
- ④ Conclusion et perspectives de recherche
 - Résumé
 - Perspectives

Perspectives



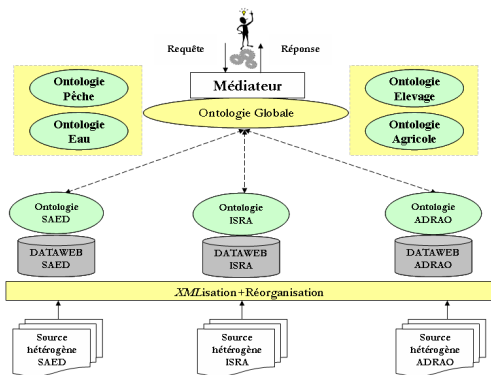
Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects THEMATIQUES
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal

Perspectives



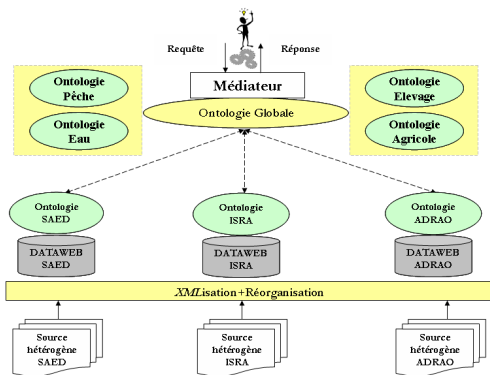
Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects **THEMATIQUES**
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal

Perspectives



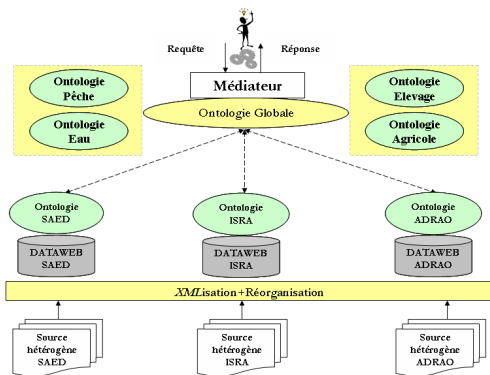
Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects THEMATIQUES
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal

Perspectives



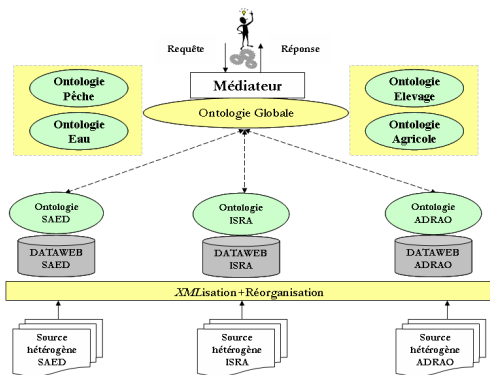
Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects THEMATIQUES
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal

Perspectives



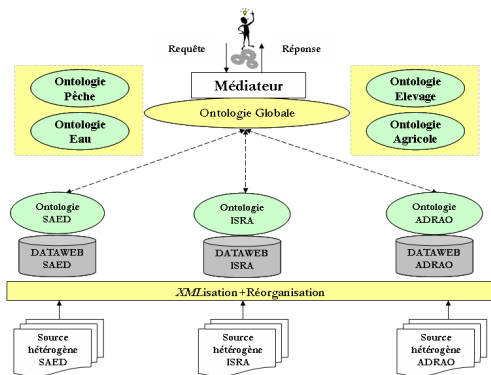
Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects THEMATIQUES
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal

Perspectives



Perspectives de recherche

- Travailler sur des volumes de données plus importants
- Intégration des aspects THEMATIQUES
- Construction de l'ontologie globale et du système de requête
- Concepts non agrégés ?
- Validation l'approche(en cours)

Perspectives de recherche

Problématique de la maintenance évolutive des ontologies du SIC-Sénégal