Towards an operator for merging taxonomies

Amélie Cordier,[1] Jean Lieber,[234] Julien Stevenot[234]

[1] Université de Lyon 1, CNRS, LIRIS, UMR5205, amelie.cordier@liris.cnrs.fr
[2] Université de Lorraine, LORIA, UMR 7503 — Vandœuvre-lès-Nancy, F-54506, France, jean.lieber@loria.fr, julien.stevenot@loria.fr
[3] CNRS, LORIA, UMR 7503 — Vandœuvre-lès-Nancy, F-54506, France
[4] Inria — Villers-lès-Nancy, F-54602, France

Taaable (http://taaable.fr)
WikiTaaable (http://wikitaaable.loria.fr)
Kolflow project (http://kolflow.univ-nantes.fr)

Outline of the talk

- Context and motivation
- Merging taxonomies
- Conclusion and future work

Context and motivation

Taaable and WikiTaaable

`http://taaable.fr`

`http://wikitaaable.loria.fr`

- Taaable: a CBR system that reuses a cooking recipe base
- WikiTaaable: a semantic wiki for the Taaable knowledge base including a taxonomical domain ontology

DSMW

- MW = MediaWiki, a wiki engine
- SMW = Semantic MW, a semantic wiki engine
- DSMW = Distributed SMW
  - Several WikiTaaables

Kolflow http://kolflow.univ-nantes.fr

- Man-machine collaboration in continuous knowledge construction flows

Kolflow http://kolflow.univ-nantes.fr

- Man-machine collaboration in continuous knowledge construction flows
- Merging the contents of two semantic wikis

Kolflow http://kolflow.univ-nantes.fr

- Man-machine collaboration in continuous knowledge construction flows
- Merging the contents of two semantic wikis
  - The textual parts

- Man-machine collaboration in continuous knowledge construction flows
- Merging the contents of two semantic wikis
  - The textual parts
  - The knowledge parts

- Man-machine collaboration in continuous knowledge construction flows
- Merging the contents of two semantic wikis
  - The textual parts
  - The knowledge parts
- Often, the two semantic wikis come from another one, so they are quite similar

Knowledge representation in a semantic wiki: mainly class-superclass relations

Knowledge representation in a semantic wiki: mainly class-superclass relations

$$\texttt{Melon} \sqsubseteq \texttt{Fruit}$$

Knowledge representation in a semantic wiki: mainly class-superclass relations

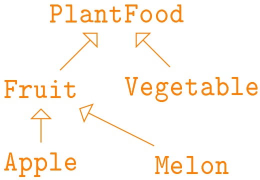
$$\texttt{Melon} \sqsubseteq \texttt{Fruit}$$

$$\forall x \quad \texttt{Melon}(x) \Rightarrow \texttt{Fruit}(x)$$

Merging taxonomies

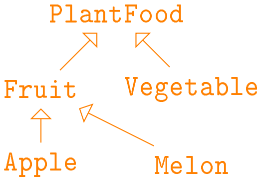
Taxonomy language

- $\mathcal{L}_{\mathcal{T}}$: language of taxonomies
  A formula of $\mathcal{L}_{\mathcal{T}}$: $A \sqsubseteq B$
  Deductive inferences based on the transitivity of $\sqsubseteq$



Taxonomy language

- $\mathcal{L}_{\mathcal{T}}$: language of taxonomies
  A formula of $\mathcal{L}_{\mathcal{T}}$: $A \sqsubseteq B$
  Deductive inferences based on the transitivity of $\sqsubseteq$
- A taxonomy ψ: a finite set of formulas of $\mathcal{L}_{\mathcal{T}}$

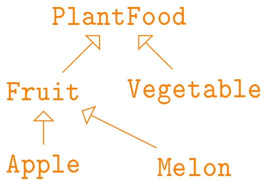


Taxonomy language

- $\mathcal{L}_{\mathcal{T}}$: language of taxonomies
  A formula of $\mathcal{L}_{\mathcal{T}}$: $A \sqsubseteq B$
  Deductive inferences based on the transitivity of $\sqsubseteq$
- A taxonomy ψ: a finite set of formulas of $\mathcal{L}_{\mathcal{T}}$
- Example:

$$\psi = \left\{ \begin{array}{ll} \texttt{Apple} \sqsubseteq \texttt{Fruit}, & \texttt{Melon} \sqsubseteq \texttt{Fruit}, \\ \texttt{Fruit} \sqsubseteq \texttt{PlantFood}, & \texttt{Vegetable} \sqsubseteq \texttt{PlantFood} \end{array} \right\}$$

$$\mathcal{V}(\psi) = \{\texttt{Apple}, \texttt{Fruit}, \texttt{Melon}, \texttt{PlantFood}, \texttt{Vegetable}\}$$



Merging two taxonomies, what does it mean?

- Usual intuition of merging ψ_1 and ψ_2:
  minimally modify ψ_1 and ψ_2 into ψ_1' and ψ_2' so that their *conjunction* is consistent

$$\Delta(\{\psi_1, \psi_2\}) = \psi_1' \wedge \psi_2'$$

Merging two taxonomies, what does it mean?

- Usual intuition of merging ψ_1 and ψ_2:
  minimally modify ψ_1 and ψ_2 into ψ_1' and ψ_2' so that their *conjunction* is consistent

$$\Delta(\{\psi_1, \psi_2\}) = \psi_1' \wedge \psi_2'$$

- Usual notion of conjunction of two knowledge bases:
  union of their axioms

Merging two taxonomies, what does it mean?

- Usual intuition of merging ψ_1 and ψ_2:
  minimally modify ψ_1 and ψ_2 into ψ_1' and ψ_2' so that their *conjunction* is consistent

$$\Delta(\{\psi_1, \psi_2\}) = \psi_1' \wedge \psi_2'$$

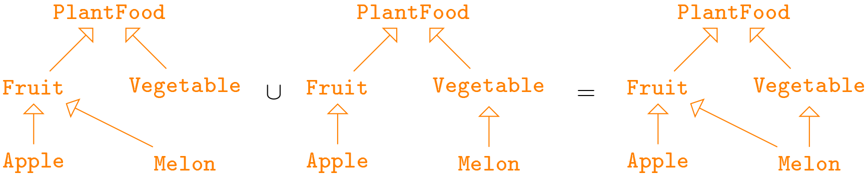
- Usual notion of conjunction of two knowledge bases:
  union of their axioms
- Problem: the union of two taxonomies is always consistent

Merging two taxonomies, what does it mean?

- Usual intuition of merging ψ_1 and ψ_2:
  minimally modify ψ_1 and ψ_2 into ψ_1' and ψ_2' so that their *conjunction* is consistent

$$\Delta(\{\psi_1, \psi_2\}) = \psi_1' \wedge \psi_2'$$

- Usual notion of conjunction of two knowledge bases:
  union of their axioms
- Problem: the union of two taxonomies is always consistent

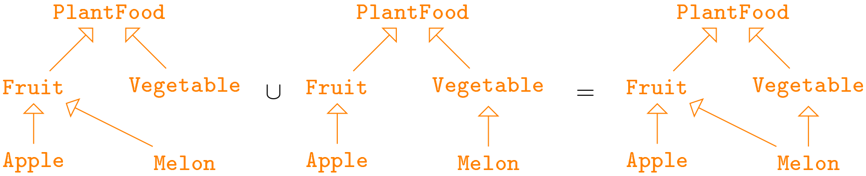


Merging two taxonomies, what does it mean?

- Usual intuition of merging ψ_1 and ψ_2:
  minimally modify ψ_1 and ψ_2 into ψ_1' and ψ_2' so that their *conjunction* is consistent

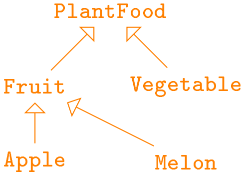
$$\Delta(\{\psi_1, \psi_2\}) = \psi_1' \wedge \psi_2'$$

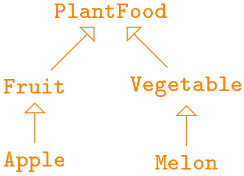
- Usual notion of conjunction of two knowledge bases:
  union of their axioms
- Problem: the union of two taxonomies is always consistent



- Another definition of $\wedge$ is proposed for taxonomies.

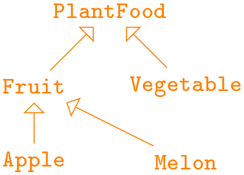
Merging two taxonomies, what does it mean?

$\psi_1 =$ 

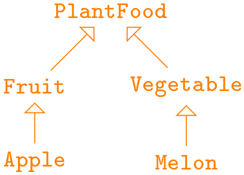
$\psi_2 =$ 

Merging two taxonomies, what does it mean?

$\psi_1 =$

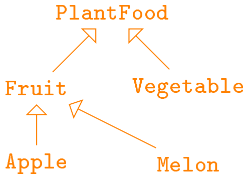


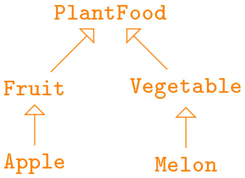
$\psi_2 =$



- $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ interpretation?

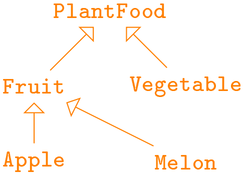
Merging two taxonomies, what does it mean?

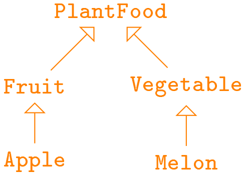
$\psi_1 =$ 

$\psi_2 =$ 

- $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ interpretation?
  1. $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ (incompleteness)

Merging two taxonomies, what does it mean?

$\psi_1 =$ 

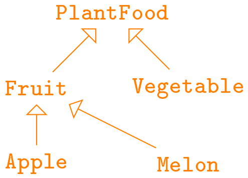
$\psi_2 =$ 

- $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ interpretation?
  1. $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ (incompleteness)
  2. $\psi_2 \models \texttt{Melon} \not\sqsubseteq \texttt{Fruit}$ (closed-world assumption, CWA)

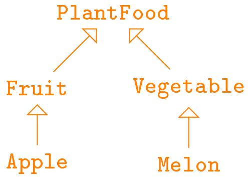
$$\frac{\psi \not\models A \sqsubseteq B}{A \not\sqsubseteq B}\,\textsf{CWA}$$

Merging two taxonomies, what does it mean?

$\psi_1 =$



$\psi_2 =$



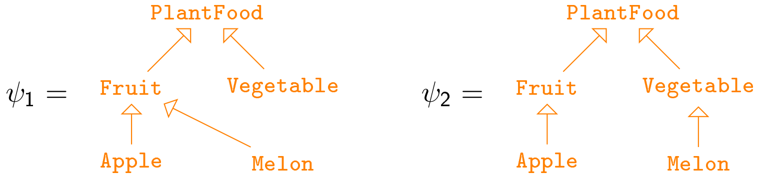
- $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ interpretation?
  1. $\psi_2 \not\models \texttt{Melon} \sqsubseteq \texttt{Fruit}$ (incompleteness)
  2. $\psi_2 \models \texttt{Melon} \not\sqsubseteq \texttt{Fruit}$ (closed-world assumption, CWA)

$$\frac{\psi \not\models A \sqsubseteq B}{A \not\sqsubseteq B}\,\text{CWA}$$

- A taxonomy ψ of $\mathcal{L}_{\mathcal{T}}$ considered under CWA:

$$\begin{aligned}\widehat{\psi} &= \{A \sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \models A \sqsubseteq B\} \\ &\cup \{A \not\sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \not\models A \sqsubseteq B\}\end{aligned}$$

Merging two taxonomies, what does it mean?



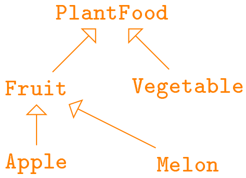
- A taxonomy ψ of $\mathcal{L}_{\mathcal{T}}$ considered under CWA:

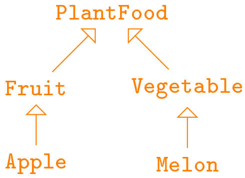
$$\widehat{\psi} = \{A \sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \models A \sqsubseteq B\} \cup \{A \not\sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \not\models A \sqsubseteq B\}$$

- Conjunction of ψ_1 and ψ_2:

$$\psi_1 \wedge \psi_2 \stackrel{def}{=} \widehat{\psi_1} \cup \widehat{\psi_2}$$

Merging two taxonomies, what does it mean?

$\psi_1 =$ 

$\psi_2 =$ 

- A taxonomy ψ of $\mathcal{L}_{\mathcal{T}}$ considered under CWA:

$$\begin{aligned}\widehat{\psi} = & \{A \sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \models A \sqsubseteq B\} \\ & \cup \{A \not\sqsubseteq B \mid A, B \in \mathcal{V}(\psi), \quad \psi \not\models A \sqsubseteq B\}\end{aligned}$$

- Conjunction of ψ_1 and ψ_2:

$$\psi_1 \wedge \psi_2 \stackrel{def}{=} \widehat{\psi_1} \cup \widehat{\psi_2}$$

- In the example, $\psi_1 \wedge \psi_2$ is inconsistent, since $\psi_1 \wedge \psi_2 \supseteq \{\texttt{Melon} \sqsubseteq \texttt{Fruit}, \texttt{Melon} \not\sqsubseteq \texttt{Fruit}\}$

Taxonomy language with negations: $\mathcal{L}_{\mathcal{T}}^{\neg}$

- $\widehat{\cdot} : \psi \in \mathcal{L}_{\mathcal{T}} \mapsto \widehat{\psi} \in \mathcal{L}_{\mathcal{T}}^{\neg}$

Taxonomy language with negations: $\mathcal{L}_\mathcal{T}^\neg$

- $\widehat{\cdot} : \psi \in \mathcal{L}_\mathcal{T} \mapsto \widehat{\psi} \in \mathcal{L}_\mathcal{T}^\neg$
- $\mathcal{L}_\mathcal{T}^\neg$'s formulas: $A \sqsubseteq B$ and $A \not\sqsubseteq B$

$$
\begin{array}{lll}
A \sqsubseteq B & \text{means} & \forall x \quad A(x) \Rightarrow B(x) \\
A \not\sqsubseteq B & \text{means} & \exists x \quad A(x) \wedge \neg B(x)
\end{array}
$$

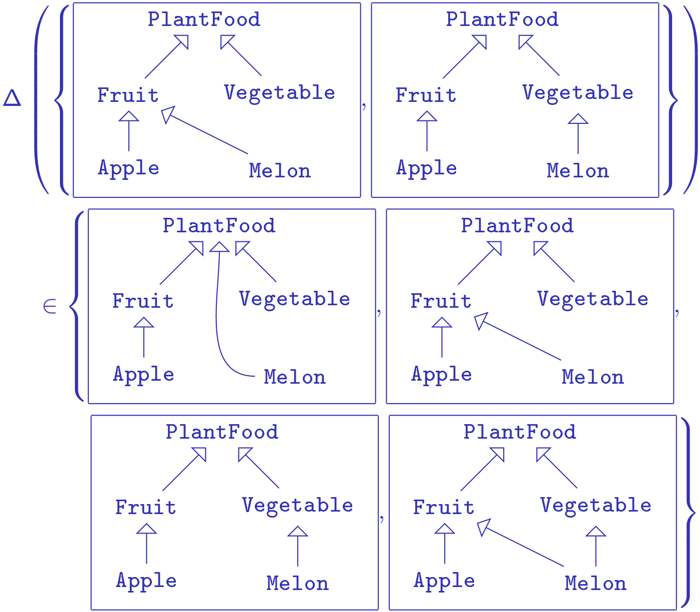
Taxonomy language with negations: $\mathcal{L}_{\mathcal{T}}^{\neg}$

- $\widehat{\cdot} : \psi \in \mathcal{L}_{\mathcal{T}} \mapsto \widehat{\psi} \in \mathcal{L}_{\mathcal{T}}^{\neg}$
- $\mathcal{L}_{\mathcal{T}}^{\neg}$'s formulas: $A \sqsubseteq B$ and $A \not\sqsubseteq B$

$$\begin{array}{llll} A \sqsubseteq B & \text{means} & \forall x & A(x) \Rightarrow B(x) \\ A \not\sqsubseteq B & \text{means} & \exists x & A(x) \wedge \neg B(x) \end{array}$$

- Remark: $\mathcal{L}_{\mathcal{T}}^{\neg}$ is not propositionnaly closed

Expected result of merging, on the example



Definition of a merging operator (1/2)

- Input: a set $\{\psi_1, \ldots, \psi_n\}$ of taxonomies
  (in practice, for the Kolflow project: $n = 2$)

Definition of a merging operator (1/2)

- Input: a set $\{\psi_1, \ldots, \psi_n\}$ of taxonomies
  (in practice, for the Kolflow project: $n = 2$)
- Output: $\Delta(\{\psi_1, \ldots, \psi_n\})$: a taxonomy

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ *// agreement*

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ // *agreement*
3. $\delta_i := \psi_i \setminus \alpha$ for each i

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ *// agreement*
3. $\delta_i := \psi_i \setminus \alpha$ for each i
4. $\delta := \bigcup_i \delta_i$ *// disagreement*

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ *// agreement*
3. $\delta_i := \psi_i \setminus \alpha$ for each i
4. $\delta := \bigcup_i \delta_i$ *// disagreement*
5. $\texttt{Candidates} := \left\{ \Gamma \;\middle|\; \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ *// agreement*
3. $\delta_i := \psi_i \setminus \alpha$ for each i
4. $\delta := \bigcup_i \delta_i$ *// disagreement*
5. $\texttt{Candidates} := \left\{ \Gamma \;\middle|\; \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$
6. $\Gamma :=$ choice in $\texttt{Candidates}$ *// magic step!*

Definition of a merging operator (2/2)

1. $\psi_i := \widehat{\psi_i}$ for each i
2. $\alpha := \bigcap_i \psi_i$ *// agreement*
3. $\delta_i := \psi_i \setminus \alpha$ for each i
4. $\delta := \bigcup_i \delta_i$ *// disagreement*
5. $\texttt{Candidates} := \left\{ \Gamma \;\middle|\; \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$
6. Γ := choice in `Candidates` *// magic step!*
7. **return** deductive reduction of Γ

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]
- (A-1), (A-2) and (A-3) are satisfied by Δ

(A-1) $\Delta(E)$ is consistent.
(A-2) If $\bigwedge E$ is consistent then $\Delta(E) = \bigwedge E$.
(A-3) Irrelevance of syntax.

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]
- (A-1), (A-2) and (A-3) are satisfied by Δ

(A-1) $\Delta(E)$ is consistent.
(A-2) If $\bigwedge E$ is consistent then $\Delta(E) = \bigwedge E$.
(A-3) Irrelevance of syntax.

- (A-4), the fairness property, is *not* satisfied by Δ

(A-4) If $\psi_1 \wedge \psi_2$ is not consistent then $\Delta(\{\psi_1, \psi_2\}) \not\models \psi_1$.

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]
- (A-1), (A-2) and (A-3) are satisfied by Δ

(A-1) $\Delta(E)$ is consistent.
(A-2) If $\bigwedge E$ is consistent then $\Delta(E) = \bigwedge E$.
(A-3) Irrelevance of syntax.

- (A-4), the fairness property, is *not* satisfied by Δ

(A-4) If $\psi_1 \wedge \psi_2$ is not consistent then $\Delta(\{\psi_1, \psi_2\}) \not\models \psi_1$.

Interpretation: This is due to the lack of disjunction in $\mathcal{L}_\mathcal{T}$ that involves the necessity to make (unfair) choices.
Remark: $\nabla : \{\psi_1, \ldots, \psi_n\} \mapsto \bigvee \texttt{Candidates}$ satisfies (A-4).

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]
- (A-1), (A-2) and (A-3) are satisfied by Δ

(A-1) $\Delta(E)$ is consistent.
(A-2) If $\bigwedge E$ is consistent then $\Delta(E) = \bigwedge E$.
(A-3) Irrelevance of syntax.

- (A-4), the fairness property, is *not* satisfied by Δ

(A-4) If $\psi_1 \wedge \psi_2$ is not consistent then $\Delta(\{\psi_1, \psi_2\}) \not\models \psi_1$.

Interpretation: This is due to the lack of disjunction in $\mathcal{L}_{\mathcal{T}}$ that involves the necessity to make (unfair) choices.
Remark: $\nabla : \{\psi_1, \dots, \psi_n\} \mapsto \bigvee \texttt{Candidates}$ satisfies (A-4).

- (A-5) and (A-6) only proven for binary merging ($n = 2$): sorry!

(A-5) $\Delta(E1) \wedge \Delta(E_2) \models \Delta(E_1 \cup E_2)$
(A-6) If $\Delta(E1) \wedge \Delta(E_2)$ is consistent then $\Delta(E_1 \cup E_2) \models \Delta(E1) \wedge \Delta(E_2)$.

Properties of Δ

- (A1-6): postulates of [Konieczny and Pino Pérez, 2002]
- (A-1), (A-2) and (A-3) are satisfied by Δ

(A-1) $\Delta(E)$ is consistent.
(A-2) If $\bigwedge E$ is consistent then $\Delta(E) = \bigwedge E$.
(A-3) Irrelevance of syntax.

- (A-4), the fairness property, is *not* satisfied by Δ

(A-4) If $\psi_1 \wedge \psi_2$ is not consistent then $\Delta(\{\psi_1, \psi_2\}) \not\models \psi_1$.

Interpretation: This is due to the lack of disjunction in $\mathcal{L}_\mathcal{T}$ that involves the necessity to make (unfair) choices.

Remark: $\nabla : \{\psi_1, \ldots, \psi_n\} \mapsto \bigvee \texttt{Candidates}$ satisfies (A-4).

- (A-5) and (A-6) only proven for binary merging ($n = 2$): sorry!

(A-5) $\Delta(E1) \wedge \Delta(E_2) \models \Delta(E_1 \cup E_2)$
(A-6) If $\Delta(E1) \wedge \Delta(E_2)$ is consistent then $\Delta(E_1 \cup E_2) \models \Delta(E1) \wedge \Delta(E_2)$.

- Complexity (of a straightforward algorithm): polynomial in $|\alpha|$ + exponential in $|\delta|$

Conclusion and future work

Conclusion and future work

Conclusion — A work at its beginning on merging taxonomies (or revising a taxonomy by another taxonomy)

Conclusion and future work

Conclusion A work at its beginning on merging taxonomies (or revising a taxonomy by another taxonomy)

Future work

Conclusion and future work

Conclusion A work at its beginning on merging taxonomies (or revising a taxonomy by another taxonomy)

Future work

- More studies about the properties of the operator

Conclusion and future work

Conclusion A work at its beginning on merging taxonomies (or revising a taxonomy by another taxonomy)

Future work

- More studies about the properties of the operator
- Integrating the user in the choice process (and reusing previous choices of users)

Conclusion and future work

Conclusion A work at its beginning on merging taxonomies (or revising a taxonomy by another taxonomy)

Future work

- More studies about the properties of the operator
- Integrating the user in the choice process (and reusing previous choices of users)
- Implementation, test, optimisation