

Towards an operator for merging taxonomies

Amélie Cordier,¹ Jean Lieber,²³⁴ Julien Stevenot²³⁴

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](#)

- ▶ Man-machine collaboration in continuous knowledge construction flows

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

- ▶ 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



The screenshot shows a web browser window with the address bar containing "wikitaable.loria.fr/index.php/Category:Melon". The page header features the logo "WIKITAAABLE3CCC" and navigation links: "MAIN PAGE | ABOUT | HELP | FAQ | SPECIAL PAGES | LOG IN".

On the left side, there is a search section labeled "Find" with an input field and "Go" and "Search" buttons. Below it is a "Browse" section with a list of links: "Main page", "Recipe list", "Food Ontology", "Dish types", "Dish roles", "Origins", "Diets", and "Culinary actions".

The main content area displays "Category:Melon" with the subtitle "From Wikitaable3ccc". To the right, there are links for "Category: Fruit", "Printable version", "Disclaimers", and "Privacy policy".

Under the heading "Description", the text reads: "Melon is a name given to various members of the plant family with sweet flavoured, fleshy fruit e.g. gourds or cucurbits. Melon can be referred as a". To the right of this text is a partially visible image of a yellow melon.

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



The screenshot shows a web browser window with the address bar containing 'wikitaaable.loria.fr/index.php/Category:Melon'. The page header features the logo 'WIKITAAABLE3CCC' and navigation links: 'MAIN PAGE | ABOUT | HELP | FAQ | SPECIAL PAGES | LOG IN'. On the right side, there are links for 'Category: Fruit', 'Printable version', 'Disclaimers', and 'Privacy policy'. On the left, there is a search box with 'Find' text, 'Go' and 'Search' buttons, and a 'Browse' section with links to 'Main page', 'Recipe list', 'Food Ontology', 'Dish types', 'Dish roles', 'Origins', 'Diets', and 'Culinary actions'. The main content area displays 'Category:Melon' with the subtitle 'From Wikitaaable3ccc'. Below this is a 'Description' section stating: 'Melon is a name given to various members of the plant family with sweet flavoured, fleshy fruit e.g. gourds or cucurbits. Melon can be referred as a'. To the right of the description is a partially visible image of a yellow melon.

Melon \sqsubseteq Fruit

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



The screenshot shows a web browser window with the address bar containing 'wikitaable.loria.fr/index.php/Category:Melon'. The page header features the logo 'WIKITAAABLE3CCC' and navigation links: 'MAIN PAGE | ABOUT | HELP | FAQ | SPECIAL PAGES | LOG IN'. On the right side, there is a link for 'Category: Fruit' and other utility links: 'Printable version | Disclaimers | Privacy policy'. On the left, there is a search box with 'Go' and 'Search' buttons, and a 'Browse' section with links to 'Main page', 'Recipe list', 'Food Ontology', 'Dish types', 'Dish roles', 'Origins', 'Diets', and 'Culinary actions'. The main content area is titled 'Category:Melon' with the subtitle 'From Wikitaable3ccc'. Below this is a 'Description' section stating: 'Melon is a name given to various members of the plant family with sweet flavoured, fleshy fruit e.g. gourds or cucurbits. Melon can be referred as a'. To the right of the text is a partially visible image of a yellow melon.

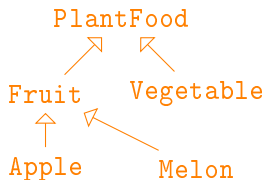
Melon \sqsubseteq Fruit

$\forall x \text{ Melon}(x) \Rightarrow \text{Fruit}(x)$

Merging taxonomies

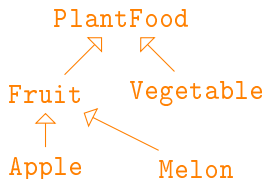
Taxonomy language

- ▶ $\mathcal{L}_{\mathcal{T}}$: 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 $\mathcal{L}_{\mathcal{T}}$

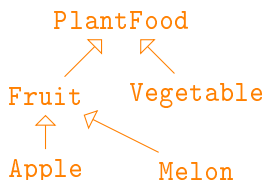


Taxonomy language

- ▶ $\mathcal{L}_{\mathcal{T}}$: 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:

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

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



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

$$\Delta(\{\psi_1, \psi_2\}) = \psi'_1 \wedge \psi'_2$$

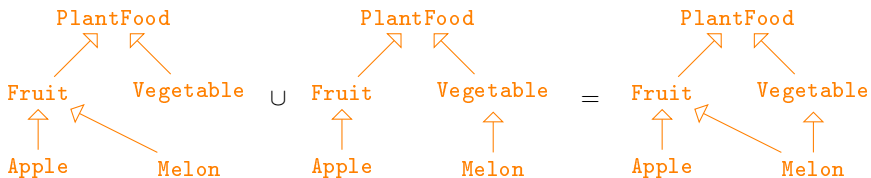
- ▶ Usual notion of conjunction of two knowledge bases:
union of their axioms
- ▶ Problem: the union of two taxonomies is always consistent

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

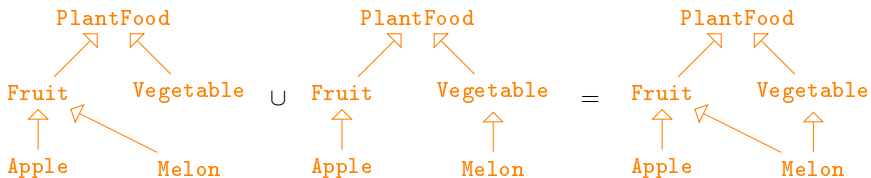


Merging two taxonomies, what does it mean?

- ▶ 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

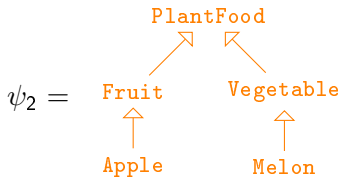
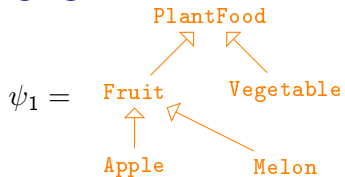
$$\Delta(\{\psi_1, \psi_2\}) = \psi'_1 \wedge \psi'_2$$

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

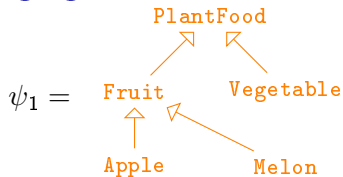


- ▶ Another definition of \wedge is proposed for taxonomies.

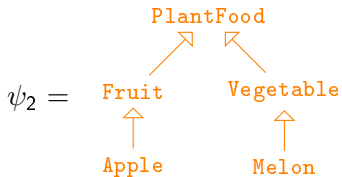
Merging two taxonomies, what does it mean?



Merging two taxonomies, what does it mean?

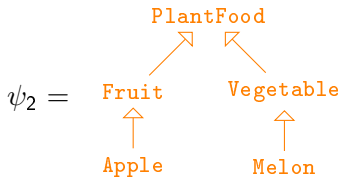
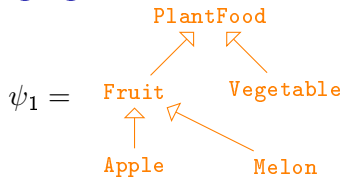


► $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$



interpretation?

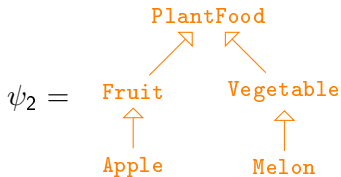
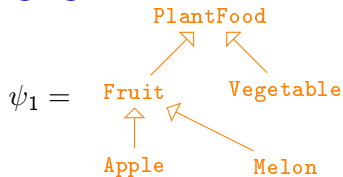
Merging two taxonomies, what does it mean?



- ▶ $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$
 1. $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$

interpretation?
(incompleteness)

Merging two taxonomies, what does it mean?

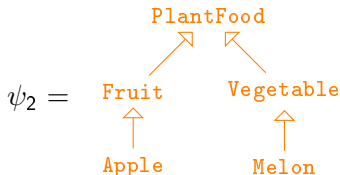
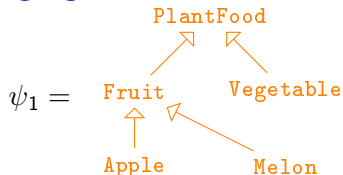


- ▶ $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$
 1. $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$
 2. $\psi_2 \models \text{Melon} \not\sqsubseteq \text{Fruit}$

interpretation?
 (incompleteness)
 (closed-world assumption, CWA)

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

Merging two taxonomies, what does it mean?



- ▶ $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$
 1. $\psi_2 \not\models \text{Melon} \sqsubseteq \text{Fruit}$
 2. $\psi_2 \models \text{Melon} \not\sqsubseteq \text{Fruit}$

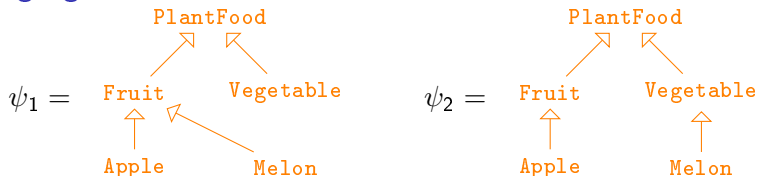
interpretation?
(incompleteness)
(closed-world assumption, CWA)

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

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

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

Merging two taxonomies, what does it mean?



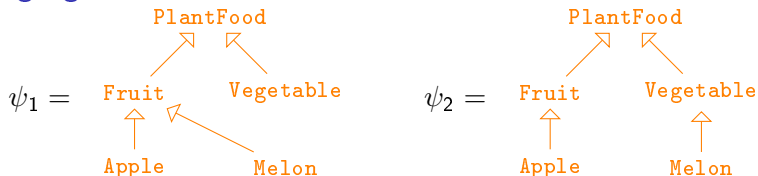
- ▶ A taxonomy ψ of $\mathcal{L}_{\mathcal{T}}$ considered under CWA:

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

- ▶ 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?



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

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

- ▶ Conjunction of ψ_1 and ψ_2 :

$$\psi_1 \wedge \psi_2 \stackrel{def}{=} \widehat{\psi}_1 \cup \widehat{\psi}_2$$

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

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

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

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

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

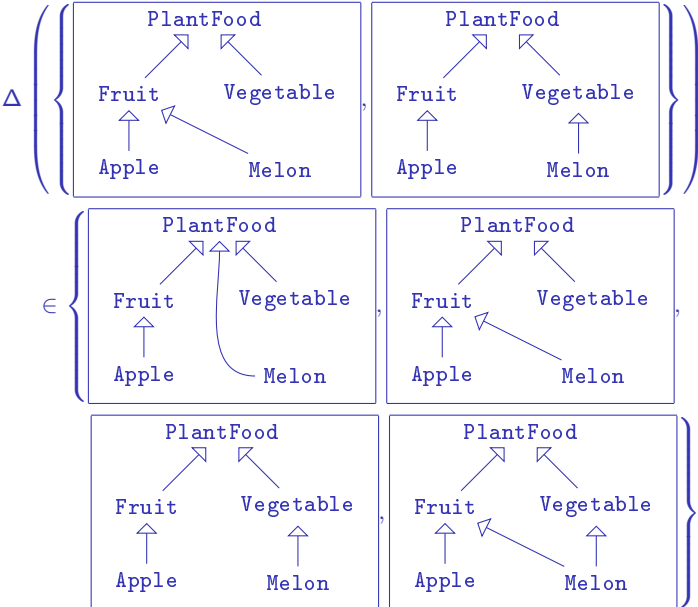
$A \sqsubseteq B$ means $\forall x \ A(x) \Rightarrow B(x)$

$A \not\sqsubseteq B$ means $\exists x \ A(x) \wedge \neg B(x)$

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

- ▶ $\hat{\cdot} : \psi \in \mathcal{L}_{\mathcal{T}} \mapsto \widehat{\psi} \in \mathcal{L}_{\mathcal{T}}^{\neg}$
- ▶ $\mathcal{L}_{\mathcal{T}}^{\neg}$'s formulas: $A \sqsubseteq B$ and $A \not\sqsubseteq B$
 - $A \sqsubseteq B$ means $\forall x \ A(x) \Rightarrow B(x)$
 - $A \not\sqsubseteq B$ means $\exists x \ A(x) \wedge \neg B(x)$
- ▶ 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 $\{\psi_1, \dots, \psi_n\}$ of taxonomies
(in practice, for the **Kolflow** project: $n = 2$)

Definition of a merging operator (1/2)

- ▶ Input: a set $\{\psi_1, \dots, \psi_n\}$ of taxonomies
(in practice, for the **Kolflow** project: $n = 2$)
- ▶ Output: $\Delta(\{\psi_1, \dots, \psi_n\})$: 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. $\alpha := \bigcap_i \psi_i$
3. $\delta_i := \psi_i \setminus \alpha$ for each 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
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. Candidates := $\left\{ \Gamma \mid \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$

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. Candidates := $\left\{ \Gamma \mid \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$
6. $\Gamma := \text{choice in Candidates}$ // 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. Candidates := $\left\{ \Gamma \mid \begin{array}{l} \alpha \subseteq \Gamma \subseteq \alpha \cup \delta \\ \Gamma \text{ is consistent} \\ \Gamma \text{ is maximal for } \subseteq \end{array} \right\}$
6. $\Gamma := \text{choice in Candidates}$ // 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: $\nabla : \{\psi_1, \dots, \psi_n\} \mapsto \bigvee$ 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: $\nabla : \{\psi_1, \dots, \psi_n\} \mapsto \bigvee$ Candidates satisfies (A-4).
- ▶ (A-5) and (A-6) only proven for binary merging ($n = 2$): sorry!
 - (A-5) $\Delta(E_1) \wedge \Delta(E_2) \models \Delta(E_1 \cup E_2)$
 - (A-6) If $\Delta(E_1) \wedge \Delta(E_2)$ is consistent then
 $\Delta(E_1 \cup E_2) \models \Delta(E_1) \wedge \Delta(E_2)$.

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

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

(A-5) $\Delta(E_1) \wedge \Delta(E_2) \models \Delta(E_1 \cup E_2)$

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

- ▶ 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