

A first draft on a tool for ontology alignment tolerance

Loïc Jeanson¹²

¹Centre François Viète
Université de Nantes

²LS2N (Laboratoire des Sciences du Numérique de Nantes)
Université de Nantes

Doctoral Symposium on Research on Online Databases in
History, April 4th-5th 2019, Leipzig

Outline

- 1 Alignment
 - What it means to align ?
 - Alignment methods
- 2 Limits of aligning ontologies
 - What the problem consists of
 - What makes the problem not so big
- 3 Draft proposal
 - The conceptual approach
 - The implementation plan

Outline

1 Alignment

- What it means to align ?
- Alignment methods

2 Limits of aligning ontologies

- What the problem consists of
- What makes the problem not so big

3 Draft proposal

- The conceptual approach
- The implementation plan

Defining ontology alignment (1/3)

- Ontology mapping finds correspondences between entities of multiple ontologies.

Given ontologies, $O_1 = (S_1, A_1)$ and $O_2 = (S_2, A_2)$, ontological mapping is a morphism $f : S_1 \rightarrow S_2$, such that $A_2 \models f(A_1)$ where all correspondences that satisfy O_2 's axioms also satisfy O_1 's "translated axiom". A correspondence is a function that assigns symbol of one vocabulary to the symbol of another vocabulary. A set of such correspondences between a pair of ontologies is an alignment. Mapping is a directed alignment.

Partial ontology mapping, is defined as having a sub-ontology $O'_1 = (S'_1, A'_1)$ where $S'_1 \subseteq S_1, A'_1 \subseteq A_1$ such that there is a total mapping from O'_1 to O_2 ¹

¹from: Yew, Kwang Hooi & Hassan, Mohd Fadzil & M Shariff, Azmi. (2014). A Survey on Ontology Mapping Techniques. Lecture Notes in

Defining ontology alignment (2/3)

- Given two ontologies, a correspondence is a 5-tuple:
 $\langle id, e1, e2, n, R \rangle$ such that
- id is a unique identifier of the given correspondence;
 - $e1$ and $e2$ are entities (e.g., tables, XML elements, properties, classes) of the first and the second ontology, respectively;
 - n is a confidence measure (typically in the $[0, 1]$ range) holding for the correspondence between $e1$ and $e2$;
 - R is a relation (e.g., equivalence($=$), more general(\sqsupseteq), disjointness (\perp), overlapping (\sqcap) holding between $e1$ and $e2$.²

²from: Gal, A., & Shvaiko, P. (2008). Advances in ontology matching. In Advances in web semantics i (pp. 176-198). Springer, Berlin, Heidelberg.

Defining ontology alignment (3/3)

- Basically, aligning amounts to defining a distance between entities (which can be as reduced as an equality predicate) and computing the best match between ontologies, i.e., the one that minimizes the total distance (or maximizes a similarity measure). But there are many different ways to compute such a distance.³

³from: Euzenat, J., & Valtchev, P. (2004, August). Similarity-based ontology alignment in OWL-lite. In ECAI (Vol. 16, p. 333).

Terminology

Matching is the *process* of finding relationships or correspondences between entities of different ontologies.

Alignment is a *set of correspondences* between two or more (in case of multiple matching) ontologies. The alignment is the output of the matching process.

Correspondence is *the relation* supposed to hold according to a particular matching algorithm or individual, between entities of different ontologies.

Mapping is the *oriented* version of an alignment.⁴

⁴from: Euzenat, J., & Shvaiko, P. (2016). Ontology matching tutorial (v17). In Proceedings of the 13th International Conference on Concept Lattices and Their Applications (CLA-'16), Moskow, Russia. 

Outline

1 Alignment

- What it means to align ?
- Alignment methods

2 Limits of aligning ontologies

- What the problem consists of
- What makes the problem not so big

3 Draft proposal

- The conceptual approach
- The implementation plan

Various methods are in use

Automatic or semi-automatic methods have been developed, based on various principles.

- *Terminological* comparison (at the level of the label or of the description, and string-based or with lexicons)
- *Internal structure* comparison
- *External structure* comparison (taxonomical structure or extensional comparison)
- *Semantic* comparison

Various tools have been developed

- Anchor-prompt (string based - edit distance measure)
- COMA (N-gram analysis, prefix/suffix)
- Cupid (Tokenization/Lemmatization)
- OLA (Tokenization/Lemmatization, prefix/suffix)
- S-MAtch (distance computation in an external hierarchy reference)

Outline

- 1 Alignment
 - What it means to align ?
 - Alignment methods
- 2 Limits of aligning ontologies
 - What the problem consists of
 - What makes the problem not so big
- 3 Draft proposal
 - The conceptual approach
 - The implementation plan

Ontology modeling transforms the researcher's modeling

Digital tools alter the way we work

“The reason we are on a higher imaginative level is not because we have a finer imagination, but because we have better instruments. [. . .] a fresh instrument serves the same purpose as foreign travel; it shows things in unusual combinations. The gain is more than a mere addition; it is a transformation”⁵

⁵Alfred N.Whitehead, cited in Ihde Don (2009), Postphenomenology and Technoscience: The Peking University Lectures, Albany, SUNY Press, p. 5–23

Some points of view need diversity

Les Esquimaux, m'a-t-on affirmé, n'ont pas de nom générique pour désigner la glace; ils ont plusieurs mots (j'ai oublié le nombre exact, mais je crois que c'est beaucoup, quelque chose comme une douzaine) qui désignent spécifiquement les divers aspects que prend l'eau entre son état tout à fait liquide et les diverses manifestations de sa plus ou moins intense congélation.

Il est difficile, évidemment, de trouver un exemple équivalent en français ; il se peut que les Esquimaux n'aient qu'un mot pour désigner l'espace qui sépare leurs igloos alors que nous en avons au moins, dans nos villes, sept (rue, avenue, boulevard, place, cours, impasse, venelle) et les Anglais au moins vingt (street, avenue, crescent, place, road, row, lane, mews, gardens, terrace, yard, square, circus, grove, court, greens, houses, gate, ground, way, drive, walk), mais nous avons tout de même une nom ("artère", par exemple), qui les englobe tous. ⁶

⁶Perec, G. (1981). Penser/classer, *Les Esquimaux*. Le Genre humain, (2), 111-127.

Classification evolve over time

My problem, with classifications, is that they don't last; as soon as I have put some order, this order is obsolete.

As everybody, I suppose, a classification frenzy takes me; the abundance of thing to classify, the quasi-impossibility to distribute them according to real satisfying criteria, makes that I never completely manage to go through. I stop at temporary and blurry sorting, merely more efficient than the initial anarchy.

The result of it produces truly strange categories; for example, a file full of various papers on which is written "to be classified", or a drawer labeled "Urgent 1" that contains nothing, (in the drawer "Urgent 2", there are a few old photographs and in the drawer "Urgent 3" unused notebooks). Anyway, I cope.⁷

⁷Perec, G. (1981). *Penser/classer, Comment je classe*. Le Genre humain, (2), 111-127.

Classification evolve over time

- Evolution of Scientific fields and theories⁸
- Evolution of Social and sociological Categories⁹
- Evolution of Lizard classification¹⁰
- ...

⁸Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago press.

⁹Hirst, P. Q. (2010). Social Evolution and Sociological Categories (Routledge Revivals) (Vol. 5). Routledge.

¹⁰Kluge, A. G. (1967). Higher taxonomic categories of gekkonid lizards and their evolution. Bulletin of the AMNH; v. 135, article 1.

Outline

1 Alignment

- What it means to align ?
- Alignment methods

2 Limits of aligning ontologies

- What the problem consists of
- What makes the problem not so big

3 Draft proposal

- The conceptual approach
- The implementation plan

Automated tools and their work with researchers

Building the relationships, and rebuilding them

*On ne “dessine” plus un graphique une fois pour toutes.
On le “construit” et on le reconstruit (on le manipule) jusqu’au
moment où toutes les relations qu’il recèle ont été perçues¹¹*

Philological works instrumented.

*L’instrument numérique multiplie les échelles d’observation
et les artefacts signifiants mais n’épuise jamais le travail
philologique, qui, sans cesse, peut proposer de nouvelles in-
terprétations.¹²*

¹¹Bertin, J. (1973). Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes (No. BOOK). Gauthier-VillarsMouton & Cie.

¹²Teissier, P., Quantin, M., & Hervy, B. (2018). Humanités numériques et archives orales: cartographies d'une mémoire collective sur les matériaux. Cahiers François Viète, 3(4), 141-177.

Outline

- 1 Alignment
 - What it means to align ?
 - Alignment methods
- 2 Limits of aligning ontologies
 - What the problem consists of
 - What makes the problem not so big
- 3 Draft proposal
 - The conceptual approach
 - The implementation plan

The ontology alignment tolerancing

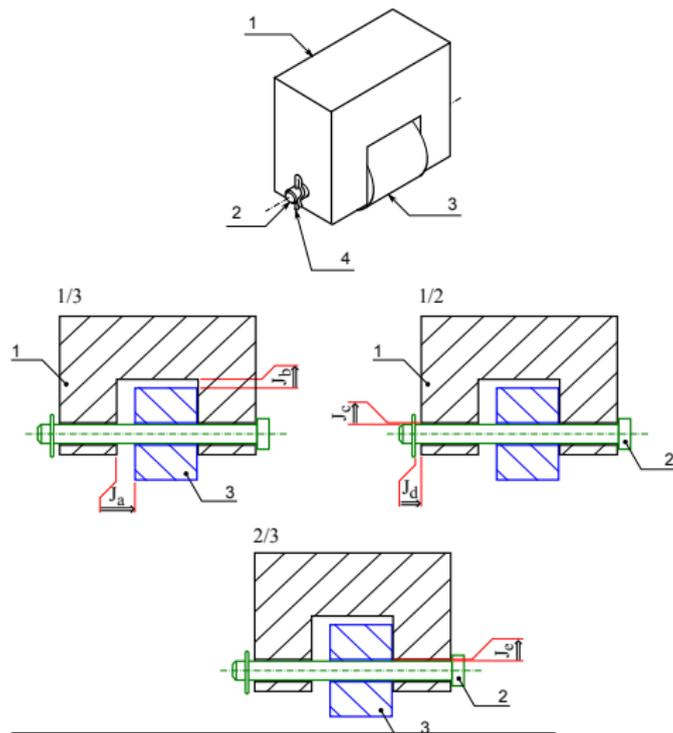
Terminology from manufacturing :

“Le jeu est l’espace laissé entre deux pièces assemblées imparfaitement” **The Backlash is the space existing between 2 parts imperfectly fitted**

“Le tolérancement, [...] définit les classes de qualité d’assemblage et fixe les règles de l’emploi du jeu mécanique”. **Tolerancing defines classes of fitting qualities and sets rules for using backlash** ¹³

¹³[https://fr.wikipedia.org/wiki/Jeu_\(m%C3%A9canique\)](https://fr.wikipedia.org/wiki/Jeu_(m%C3%A9canique)) 

Tolerancing



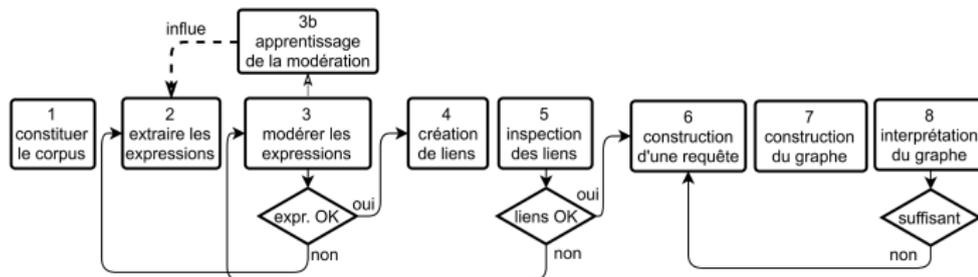
14

¹⁴from Wikimedia commons, all credit to Cdang picture under cc-by-sa 3.0

Outline

- 1 Alignment
 - What it means to align ?
 - Alignment methods
- 2 Limits of aligning ontologies
 - What the problem consists of
 - What makes the problem not so big
- 3 Draft proposal
 - The conceptual approach
 - The implementation plan

Starting from Haruspex (1/2)



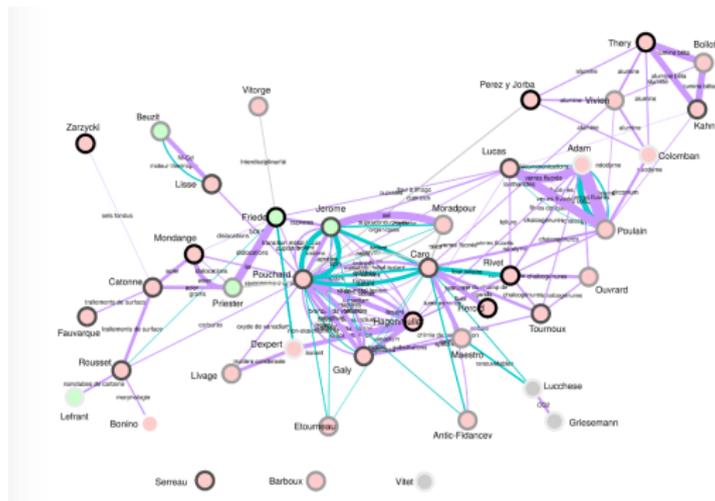
15

¹⁵from Teissier, P., Quantin, M., & Hervy, B. (2018). *Humanités numériques et archives orales: cartographies d'une mémoire collective sur les matériaux*. Cahiers François Viète, 3(4), 141-177.

Starting from Haruspex (2/2)

And in this case also a graph from the corpus grouping

16



¹⁶from Teissier, P., Quantin, M., & Hervy, B. (2018). *Humanités numériques et archives orales: cartographies d'une mémoire collective sur les matériaux*. Cahiers François Viète, 3(4), 141-177.

Using the ontology description

- Treat the elements description of the n ontologies we consider as the corpus
- Haruspex first steps
 - Tokenization / Lemmatization
 - Identification of 'relevant' words
 - Mapping of the entities in a lexical vector-space of the local corpus
- Do we need a metric? (do we need to absolute distances or only relative?)

Summary

- I think we could need a tool to *handle ontologies without aligning them* once and for all, just mapping them, keeping their differences, so that in a perspective of instrumented work, from request to request, we may specify the interesting level of tolerance. Also, with it, we could be able to track the *evolution of our ontologies* (and maybe of our domain understanding)
- I just started something in this direction, building up on Haruspex, a corpora mapping tool.
- This work is really new : comments, questions are very welcome. And any help/collaboration too!