Aligning and Merging Ontologies

Two communities with different ontologies will be able to share information when they are capable of establishing connections among their tokens in order to infer the relationship among their types. Let us develop an example taken from [1], which shows the issues one has to take into account when attempting to align the English concepts river and stream with the French concepts fleuve and rivière. According to Sowa,

> In English, size is the feature that distinguishes river from stream; in French, a fleuve is a river that flows into the sea, and a rivière is either a river or a stream that runs into another river. [1]

This explains how the concepts need to be merged. Notice that the above quote requires an agreed understanding on how to distinguish between big and small rivers, and between rivers that run into a sea or into other rivers, yielding four types of instances of 'water-flowing entities': big-into-sea, big-into-river, small-into-sea, and small-into-river.

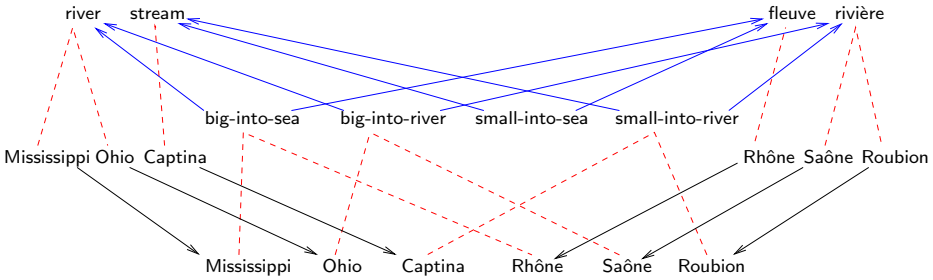

Figure 1: An agreed understanding

Figure 1 shows how both, English and French speakers, base their concepts upon this agreed understanding, although English and French speakers don't distinguish between some types of instances. For example, English speakers call both, big-into-sea and big-into-river, a river, while French speakers don't distinguish between big-into-river and small-into-river, and call both types a rivière. The agreed understanding is materialised by two maps that form the alignment. It requires the classification of particular instances of river, stream, fleuve, and rivière according with the agreed understanding, since it is this agreed way of classification which will determine how the concepts river, stream, fleuve, and rivière are going to be related to each other.

The ultimate goal is to determine the connections that link particular instances of type river or stream with particular instances of type fleuve or rivière, in a way that they respect the agreed understanding. This is done by connecting only those instances that conform to the same type according to the agreed understanding, as illustrated in Figure 2.

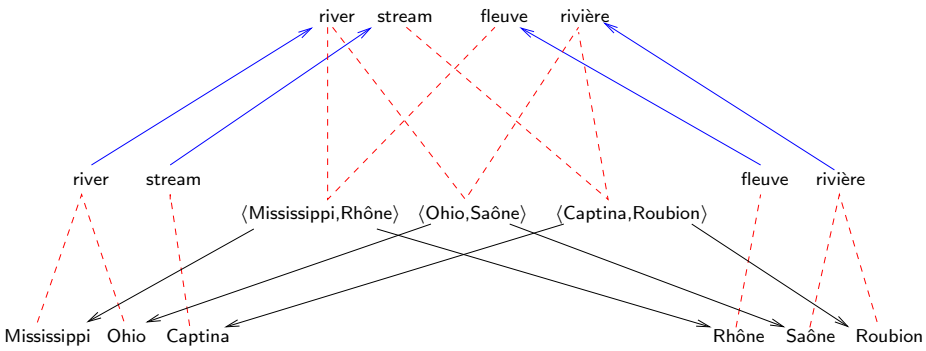

Figure 2: Aligning ontologies by means of a pair of maps

The resulting classification of connections $\langle \mathsf{Mississippi}, \mathsf{Rhône} \rangle$, $\langle \mathsf{Ohio}, \mathsf{Saône} \rangle$, and $\langle \mathsf{Captina}, \mathsf{Roubion} \rangle$ into the four concepts river, stream, fleuve, and rivière, determines a theory of how these concepts are related (e.g., that a fleuve is also a river, or that a stream is also a rivière, but not vice versa). Figure 2 shows what in channel theory is known as a an *information channel*. It captures, by means of two pairs of contra-variant functions, an existing duality between concepts and instances: Each pair consists of a map of concepts on the so called *type level* and map of instances on the so called *token level*, and pointing in the opposite direction. From a channel-theoretic perspective, Figure 2 actually illustrates us that sharing knowledge involves a flow of information that crucially depends on how the instances of different agents are connected together. The following table shows the *classification relation*, i.e., the connections as classified according to the concept types involved in the example:

	river	stream	fleuve	rivière
$\langle \mathsf{Mississippi}, \mathsf{Rhône} \rangle$	1	0	1	0
$\langle \mathsf{Ohio}, \mathsf{Saône} \rangle$	1	0	0	1
$\langle \mathsf{Captina}, \mathsf{Roubion} \rangle$	0	1	0	1

The merged set of concepts $\{\mathsf{river}, \mathsf{stream}, \mathsf{fleuve}, \mathsf{rivière}\}$ has an additional structure that we can deduce from the way the connections of instances are classified with respect to these concepts. Through techniques from formal concepts analysis, for instance, we can make such structure explicit in the form of a *concept lattice*, as shown in Figure 3. The concept hierarchy represented in this lattice depends on the choice of instances and its classification with respect to the agreed understanding. The fact that no instances were classified as of type small-into-sea was crucial in this example. Notice, also, that the resulting lattice has a node labelled with the concept $\mathsf{river} \wedge \mathsf{rivière}$, which is a formal concept that did not exist in the original vocabularies. It corresponds to the instances

of 'water-flowing entities' that, although big, flow into other rivers, like Ohio and Saône.

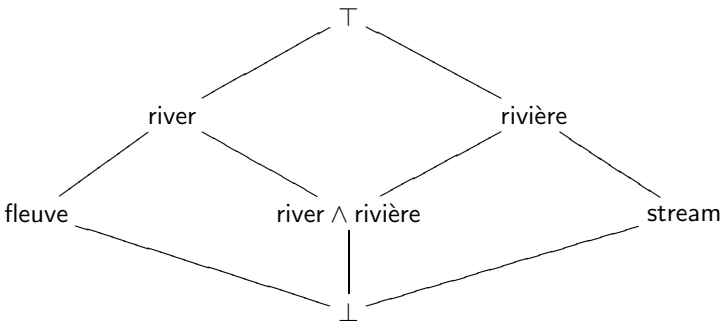

Figure 3: Concept lattice

References

[1] J. Sowa. *Knowledge Representation and Reasoning: Logical, Philosophical, and Computational Foundations.* Brooks/Cole, 2000.