

Automated information fusion under uncertainty

Matt Duckham

John Stell

Mike Worboys

NCGIA, University of Maine

David's example

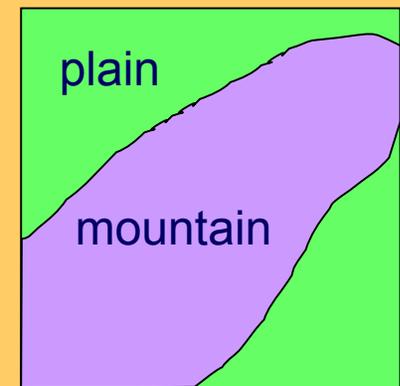
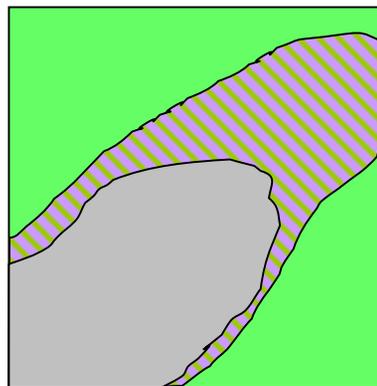
Dictionary of terms

Structure of terms



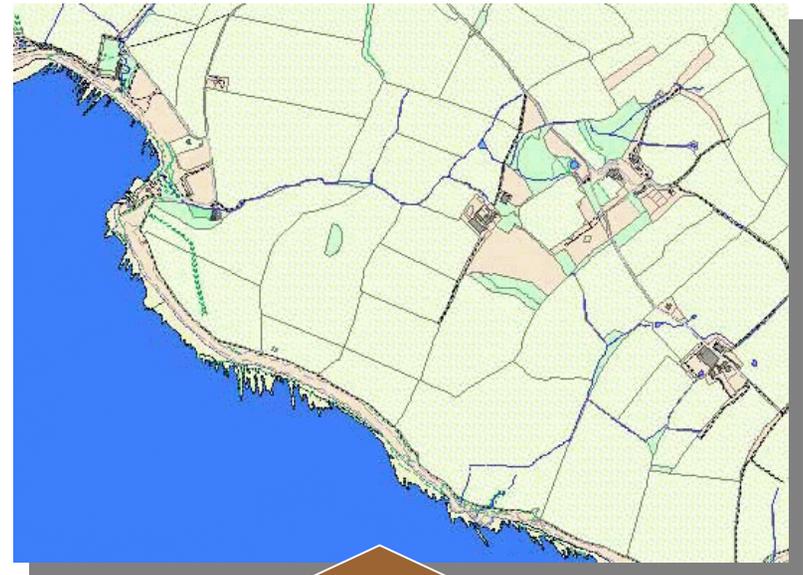
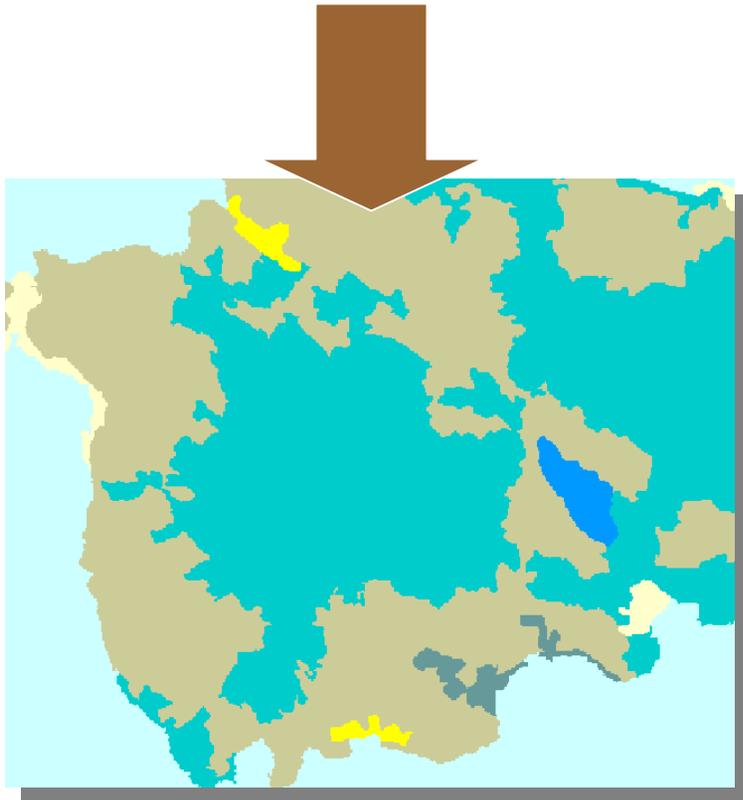
Structure of terms

Structure?



Land cover information fusion

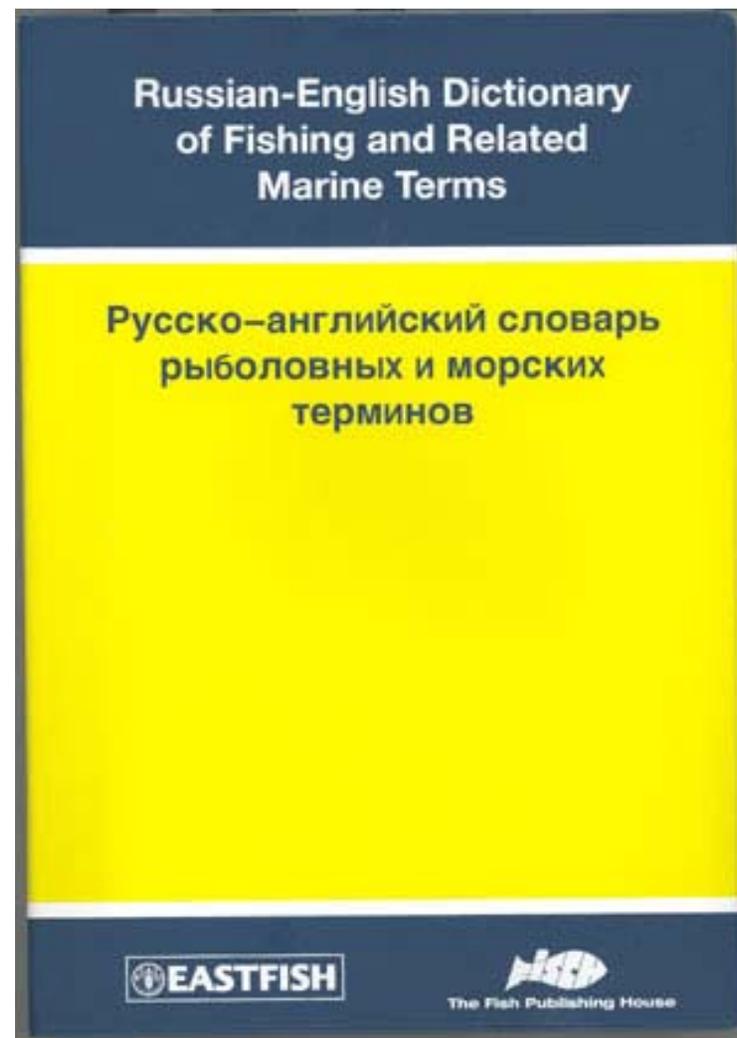
- CORINE: Fine-grained thematic, coarse-grained spatial



- OS MasterMap: coarse-grained thematic, fine grained spatial

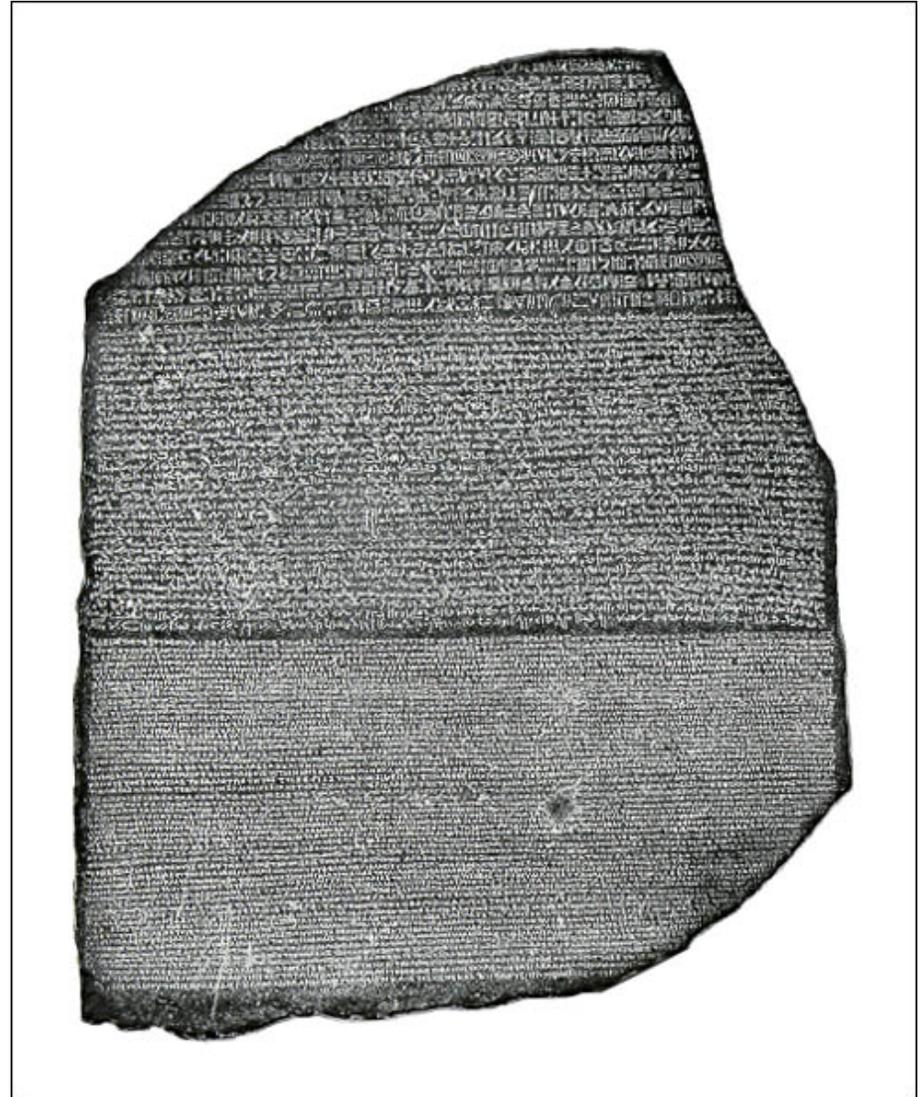
Information fusion: schema-level

- ❑ Process of integrating information from multiple sources, where the information derived is more valuable, more reliable or in some way more useful than the original information
- ❑ Existing fusion techniques largely schema-driven
- ❑ Semantics are difficult to precisely capture



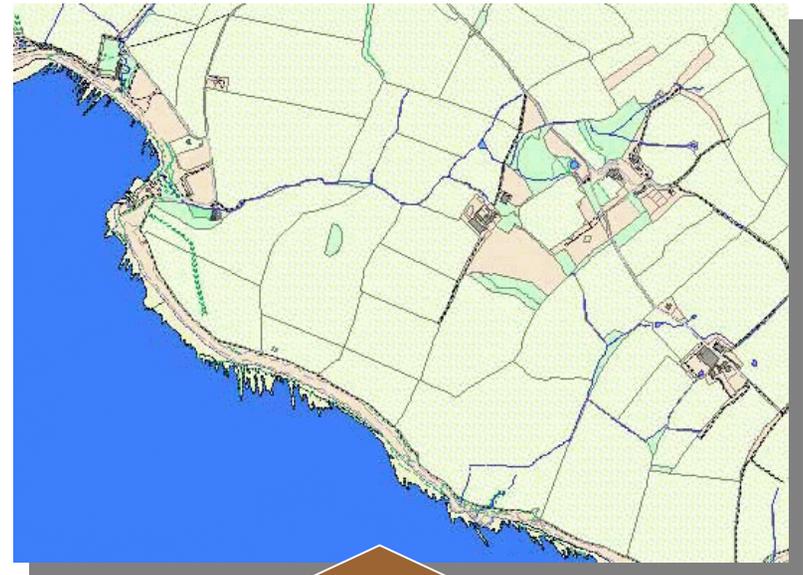
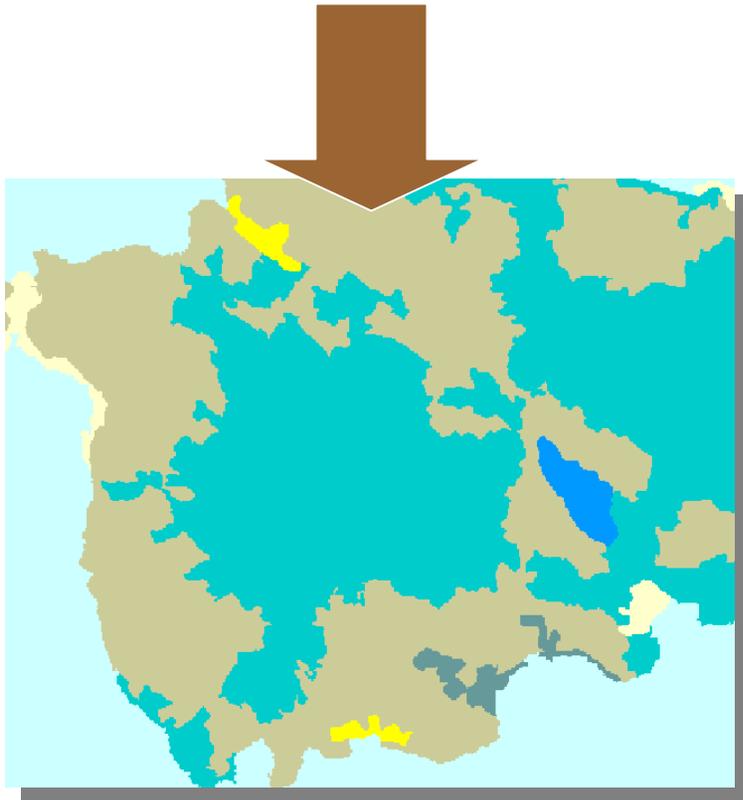
Information fusion: instance-level

- ❑ Rosetta Stone: C2nd tablet inscribed with three copies of the same decree in Egyptian (text and hieroglyphs) and Greek (text)
- ❑ C18th/19th Archaeologists used the stone to decipher Egyptian hieroglyphs
- ❑ Successful because of structure on the (instance-level) inscriptions provided a basis for (schema-level) inferences



Land cover application

- CORINE: Fine-grained thematic, coarse-grained spatial



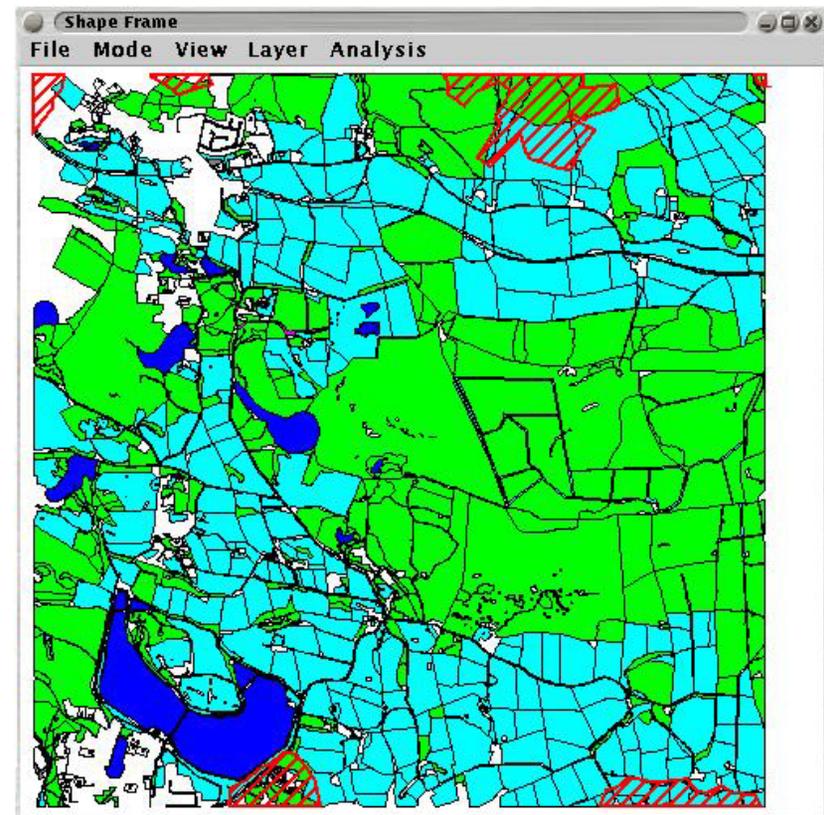
- OS MasterMap: coarse-grained thematic, fine grained spatial

Spatio-thematic layers

- An ST-layer is a distribution of elements of a thematic space over a spatial framework.
- Formally:

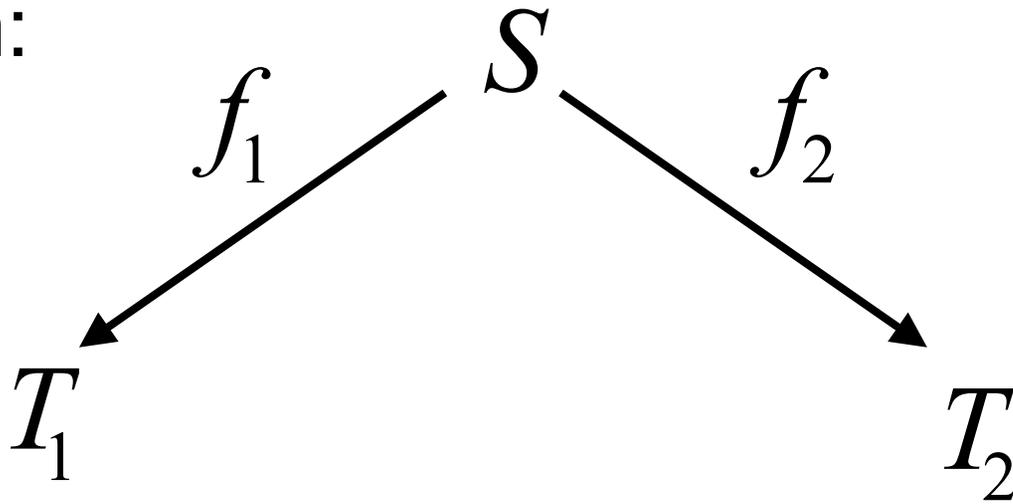
$$S \xrightarrow{f} T$$

- To:
- Assume that both layers are based on the same spatial framework S . (If not, extend each layer to the union of frameworks).

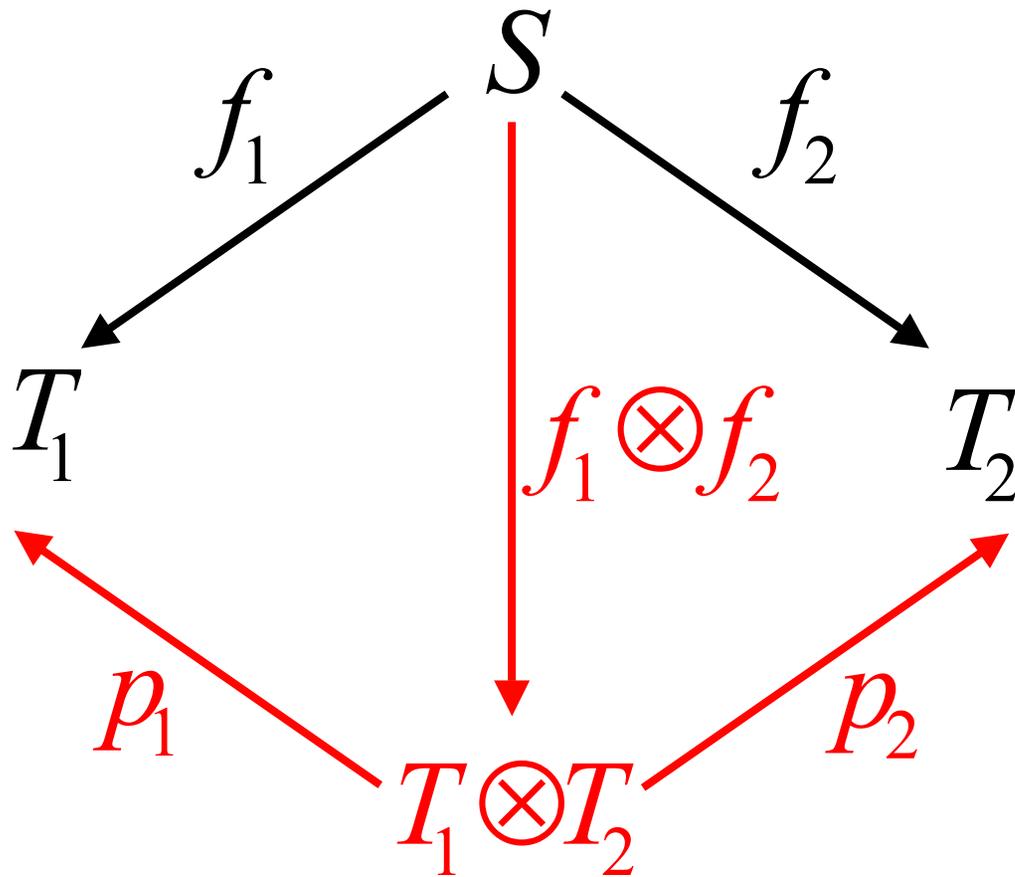


ST-layer integration

□ From:

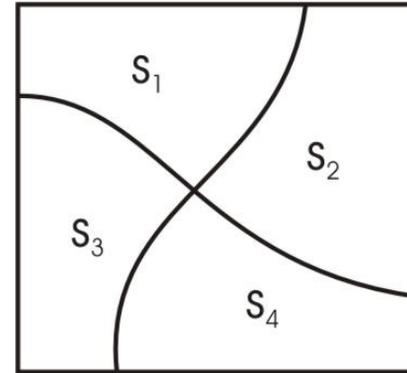


ST-layer integration

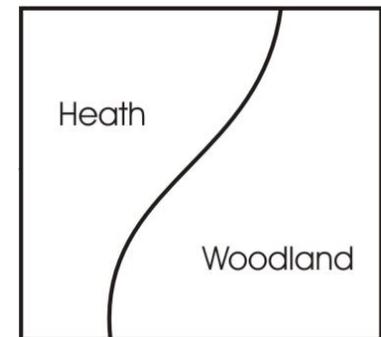
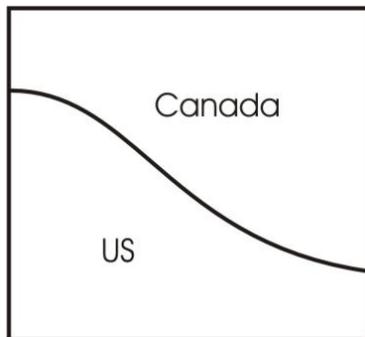


Example 1: Unstructured thematic space

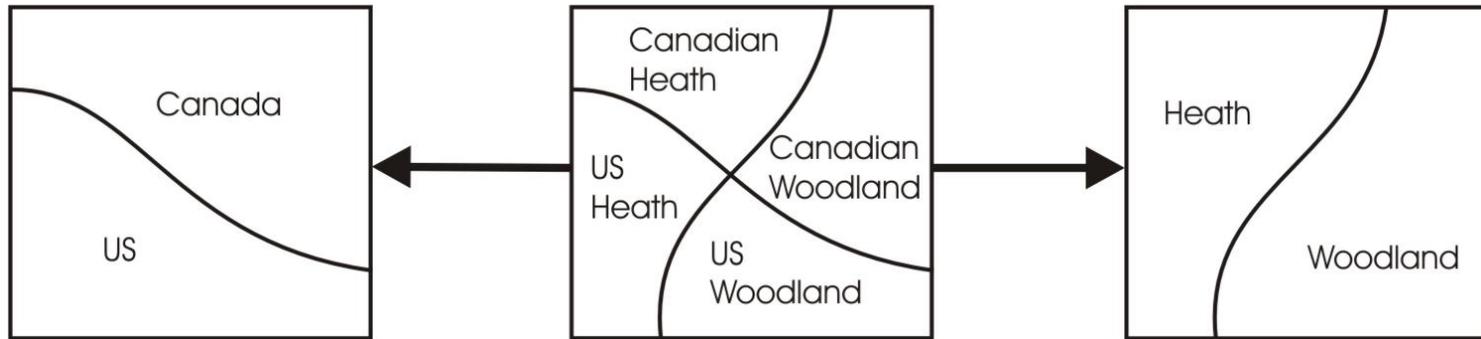
- Assume spatial framework:



- Assume ST-layers:



Integrated ST-layer



$$T_1 = \{u, c\}$$

$$T_2 = \{h, w\}$$

$$T_1 \otimes T_2 = \{(u, h), (u, w), (c, h), (c, w)\}$$

$$f_1 \otimes f_2 : s \mapsto (sf_1, sf_2)$$

Example 2: Partitioned thematic space

- Suppose we have a collection of landcover types:

c conifer woodland	b	broad-leaved woodland
n natural grassland	m	moorland
h heathland	o	orchard
p pasture	a	arable land

- Suppose two thematic spaces are given as:

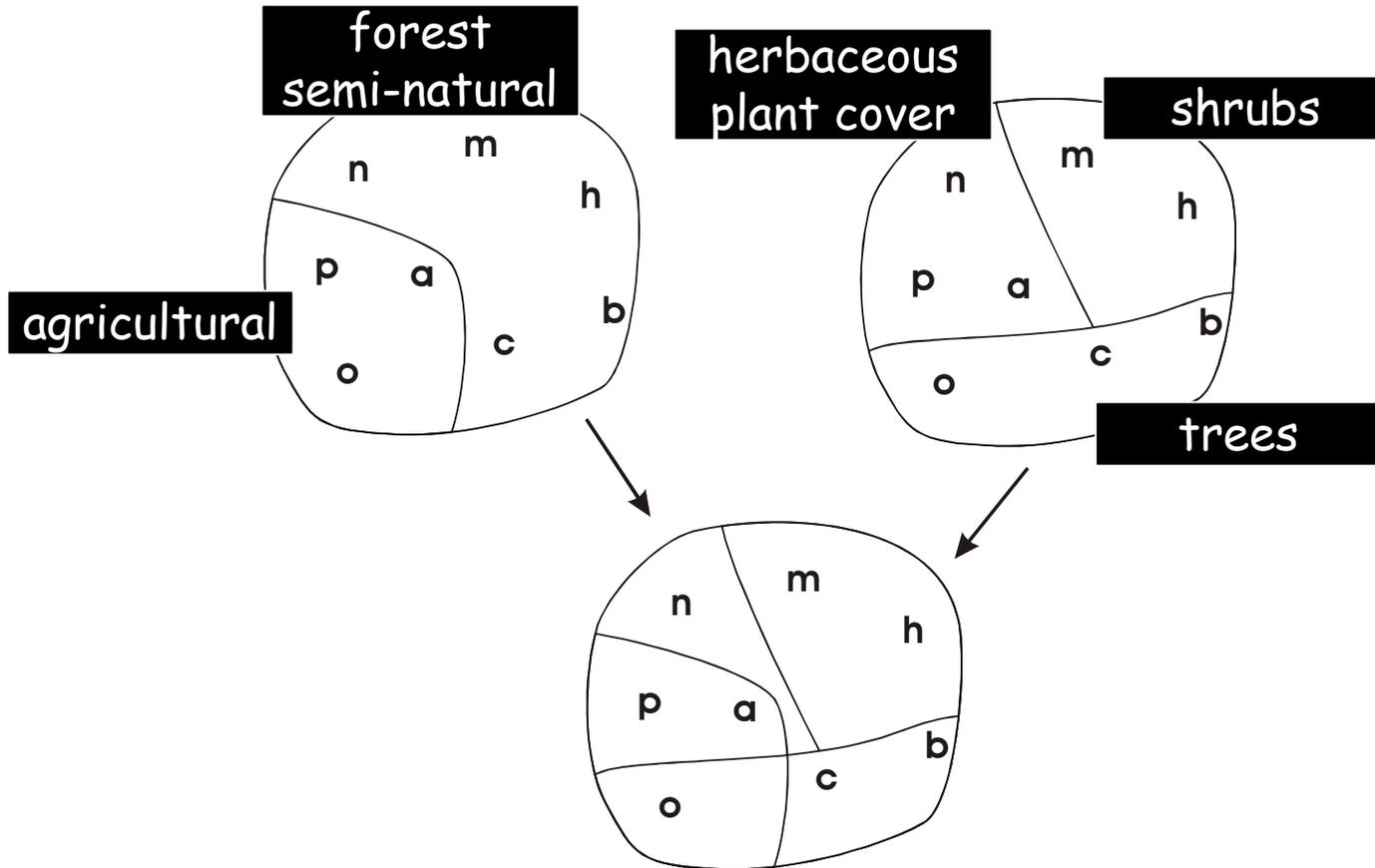
$$T_1 = \{\{p, a, o\}, \{c, b, n, m, h\}\}$$

$$T_2 = \{\{p, a, n\}, \{m, h\}, \{c, b, o\}\}$$

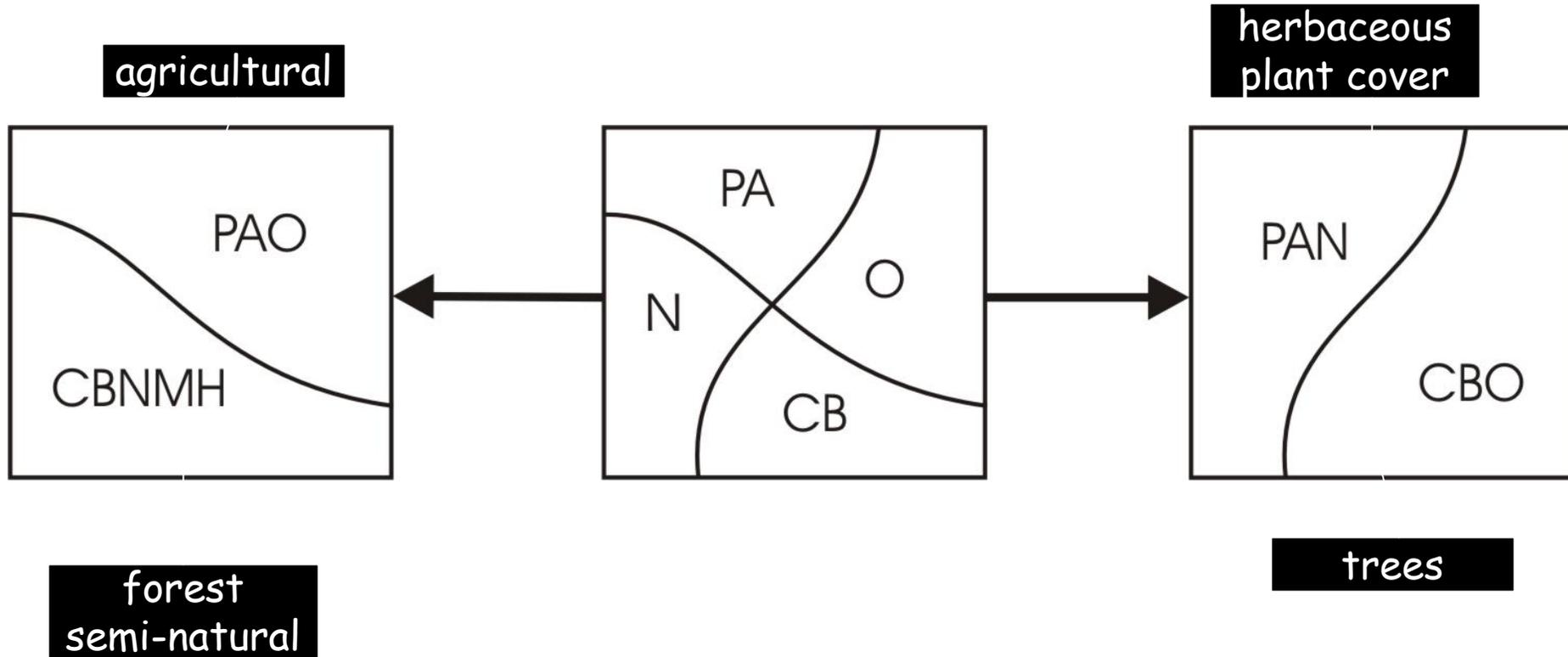
The integrated thematic space will be:

$$T_1 \otimes T_2 = \{\{p, a\}, \{o\}, \{c, b\}, \{n\}, \{m, h\}\}$$

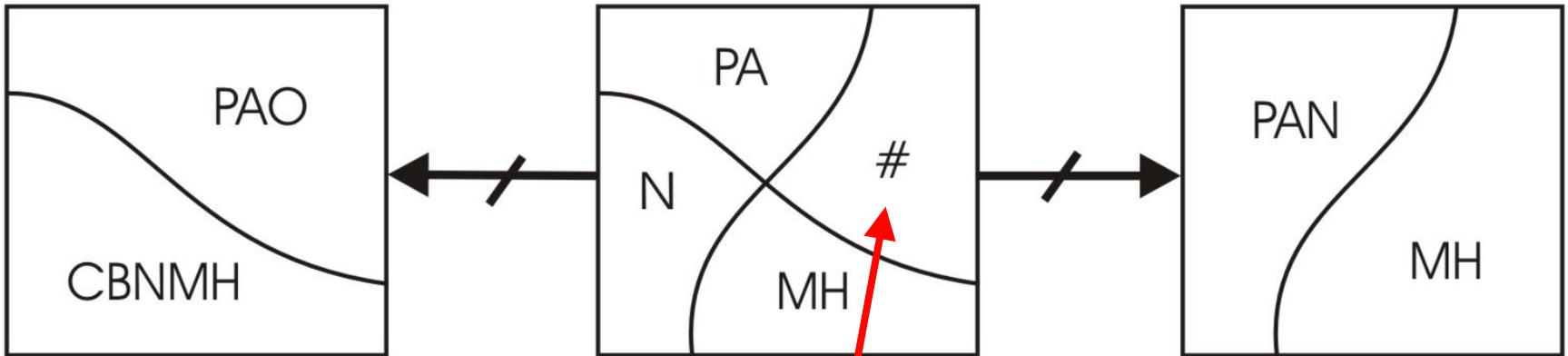
Thematic product



Spatio-thematic product



Spatio-thematic product



An incompatible region

Formally

Suppose that there is an underlying space U of 'atomic' themes, and that T_1 and T_2 are partitions of U .

Suppose f_1 and f_2 are spatio-thematic layers:

$$f_1 : S \rightarrow T_1$$

$$f_2 : S \rightarrow T_2$$

Then

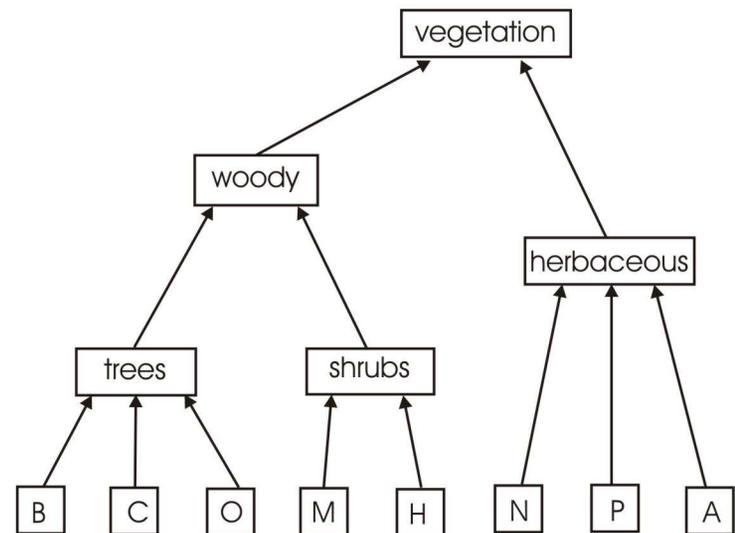
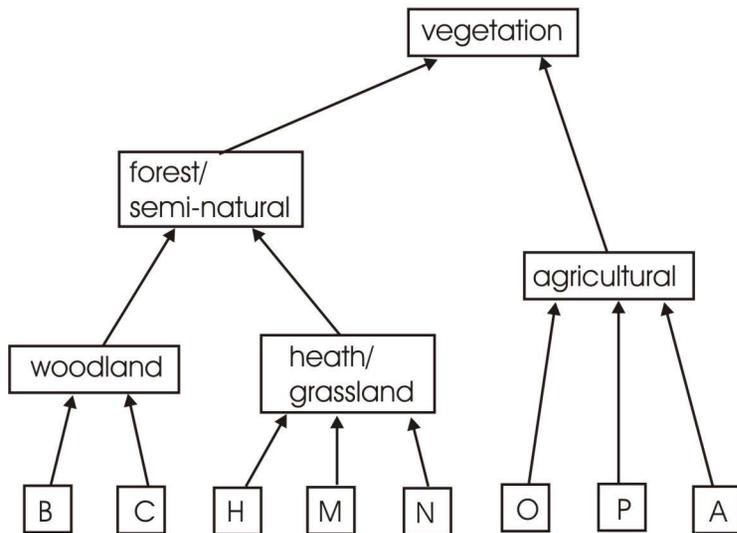
$$T_1 \otimes T_2 = \{t_1 \cap t_2 \mid t_1 \in T_1, t_2 \in T_2, t_1 \cap t_2 \neq \emptyset\}$$

$$f_1 \otimes f_2 : s \mapsto sf_1 \cap sf_2$$

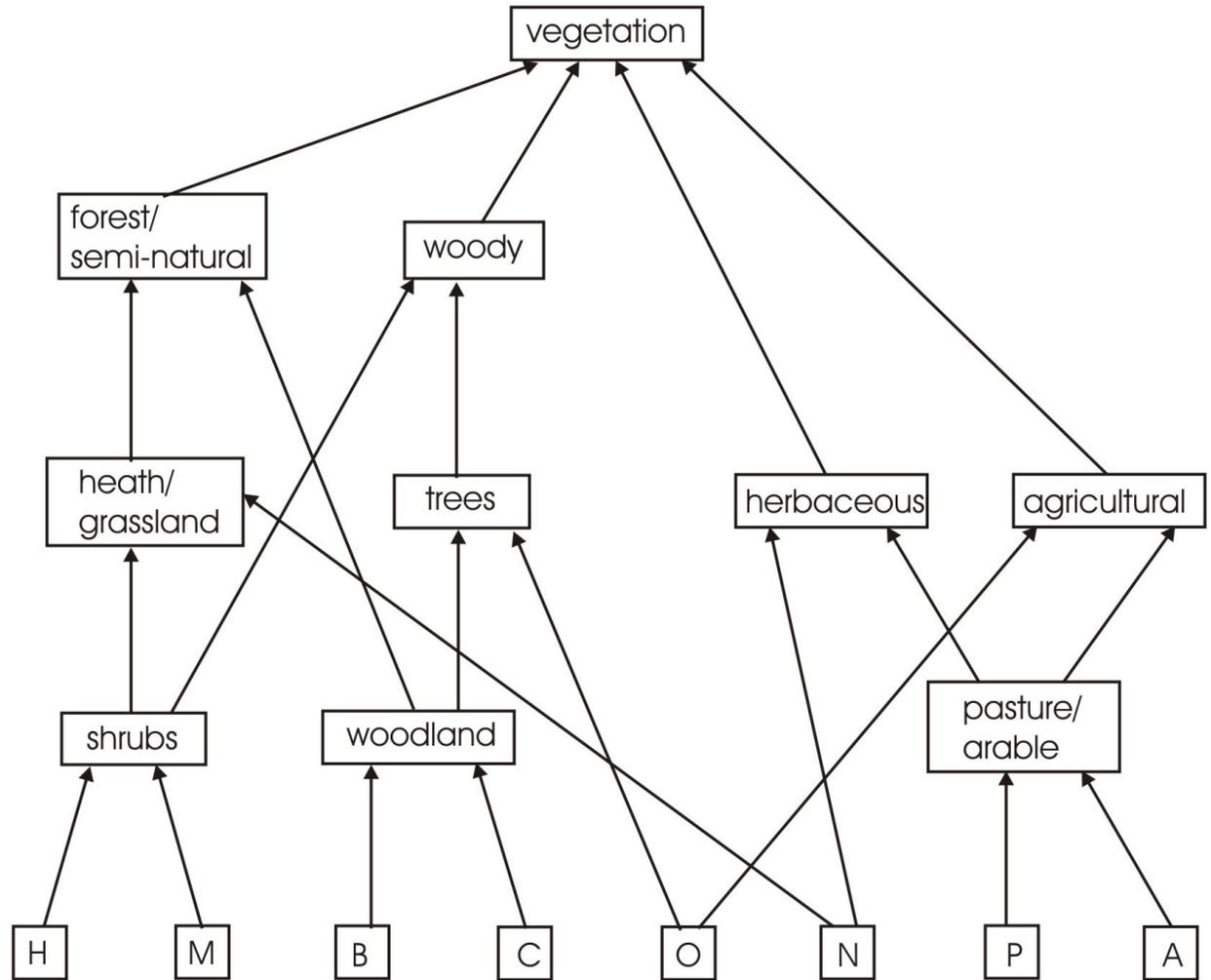
Notice that $f_1 \otimes f_2$ may not be defined, the layers may not be compatible.

**... but we are
interested in thematic
hierarchies.**

From this ...



... to this.



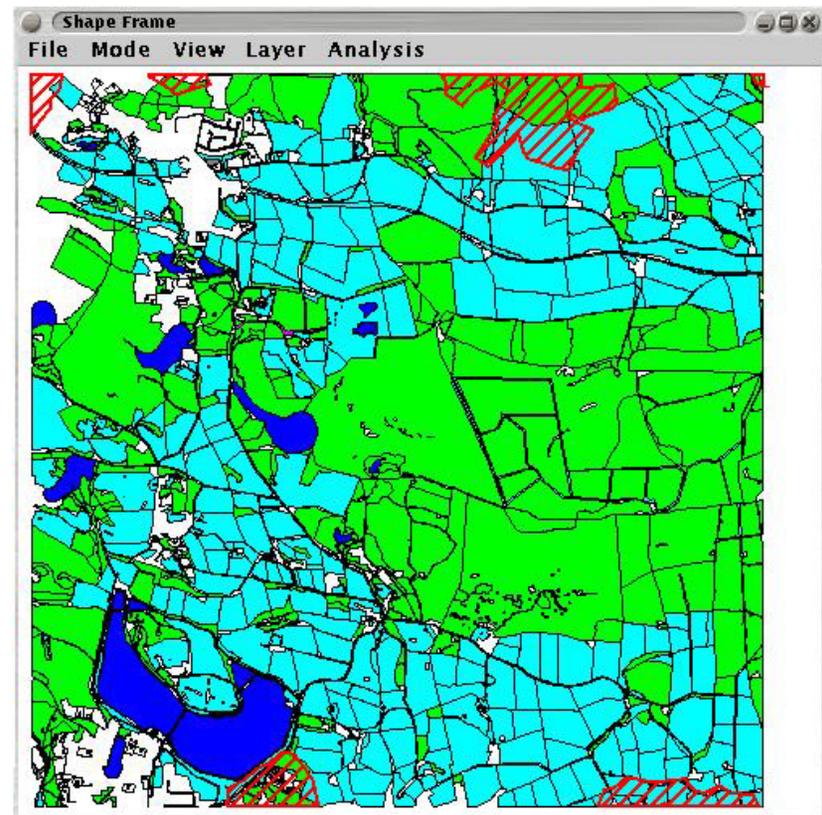
Layer or instantiation?

- An ST-layer is a distribution of elements of a thematic space over a spatial framework.

$$S \xrightarrow{f} T$$

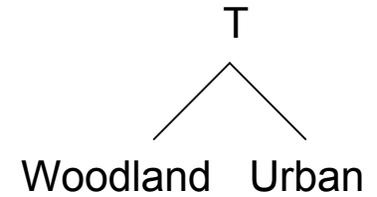
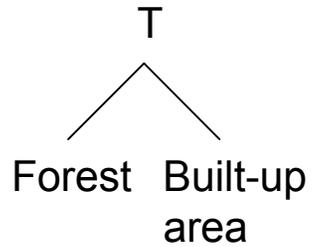
- A spatial instantiation of a classification is an association with each theme a part of the space.

$$T \xrightarrow{g} R$$

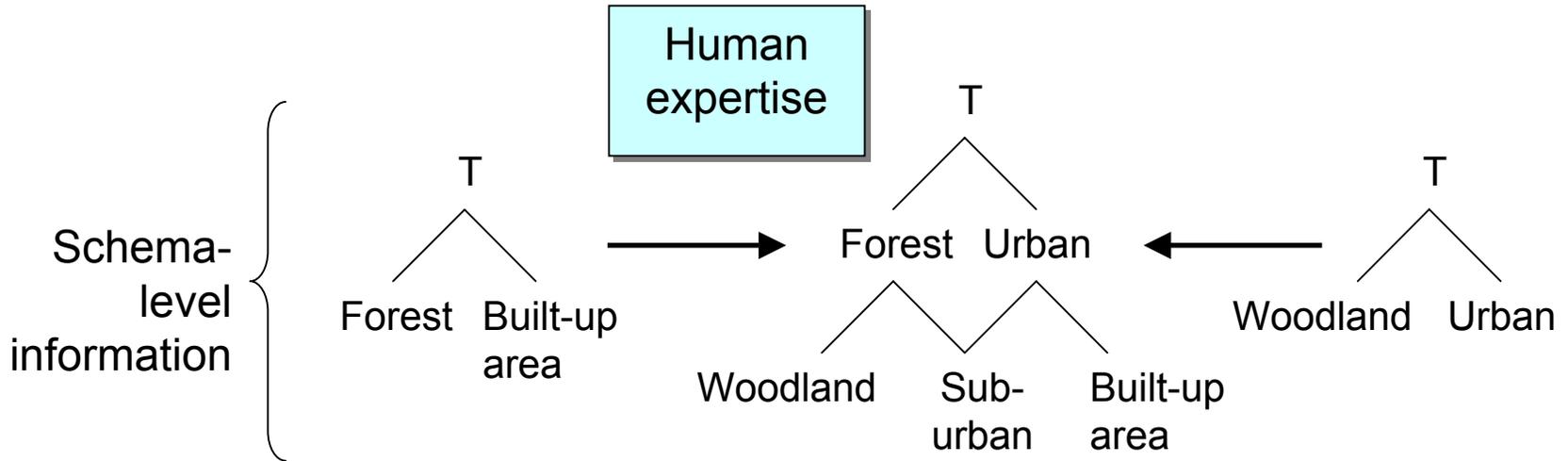


Simple example: schema-level

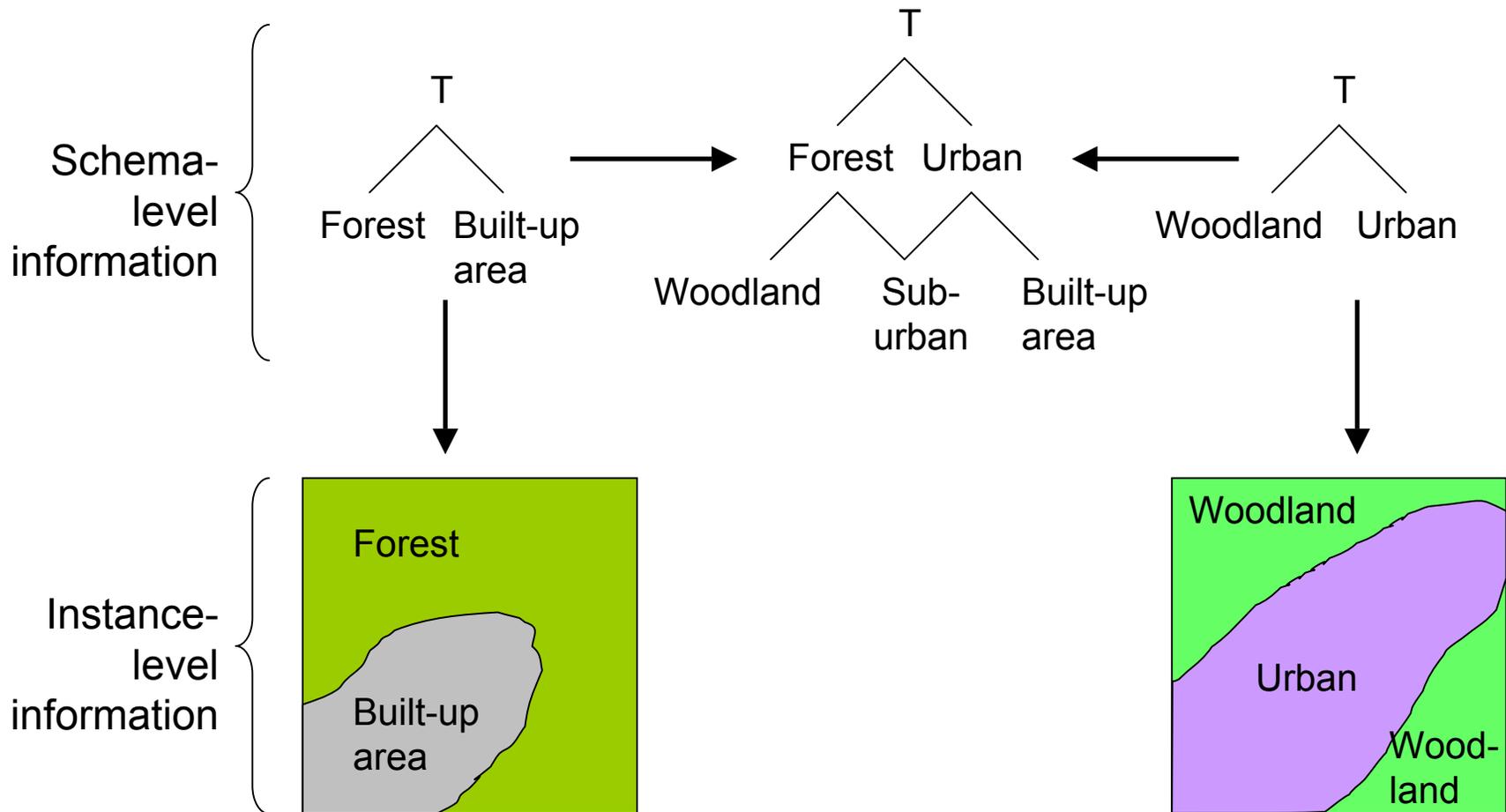
Schema-level information



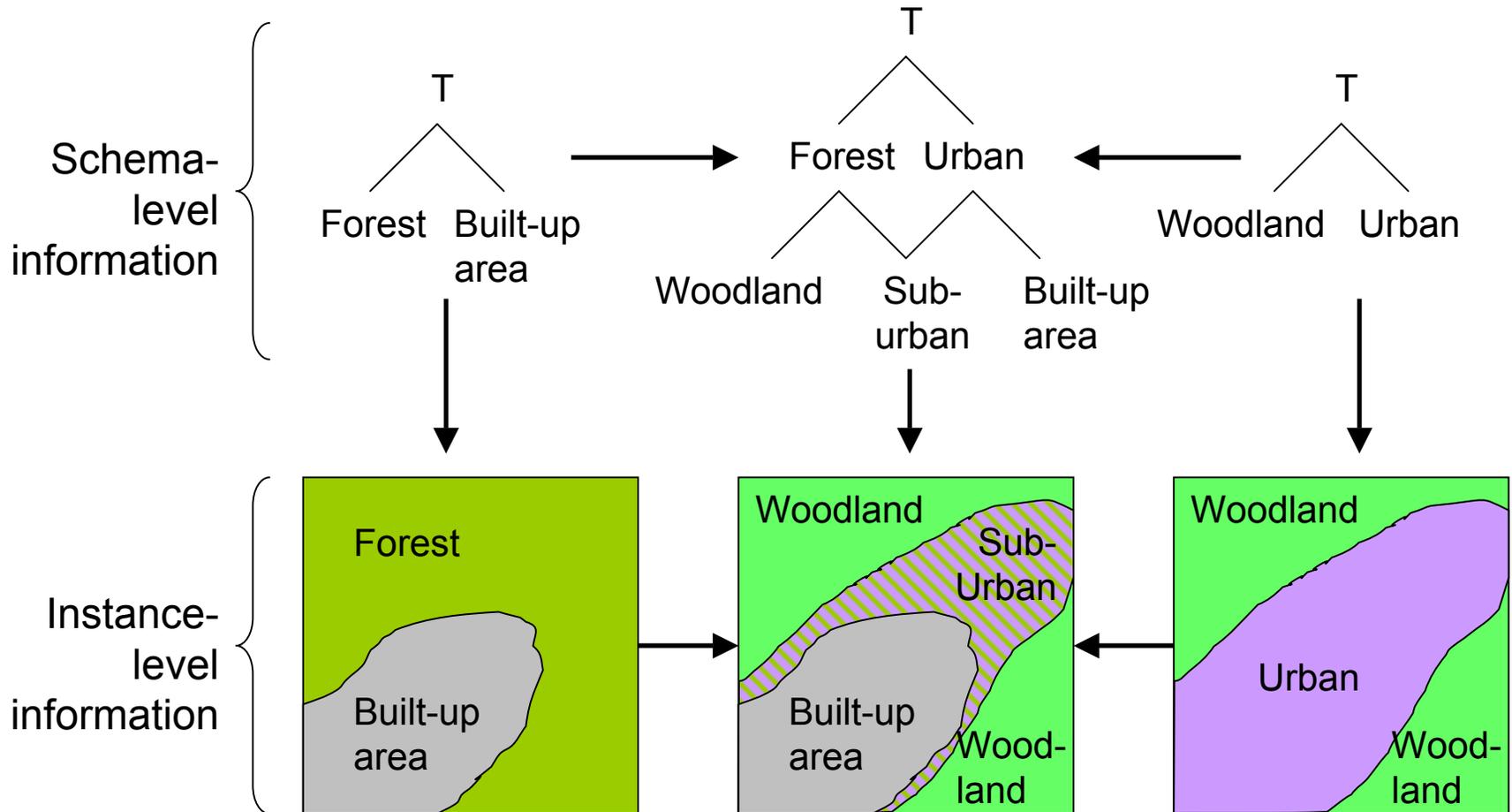
Simple example: schema-level



Simple example: schema-level



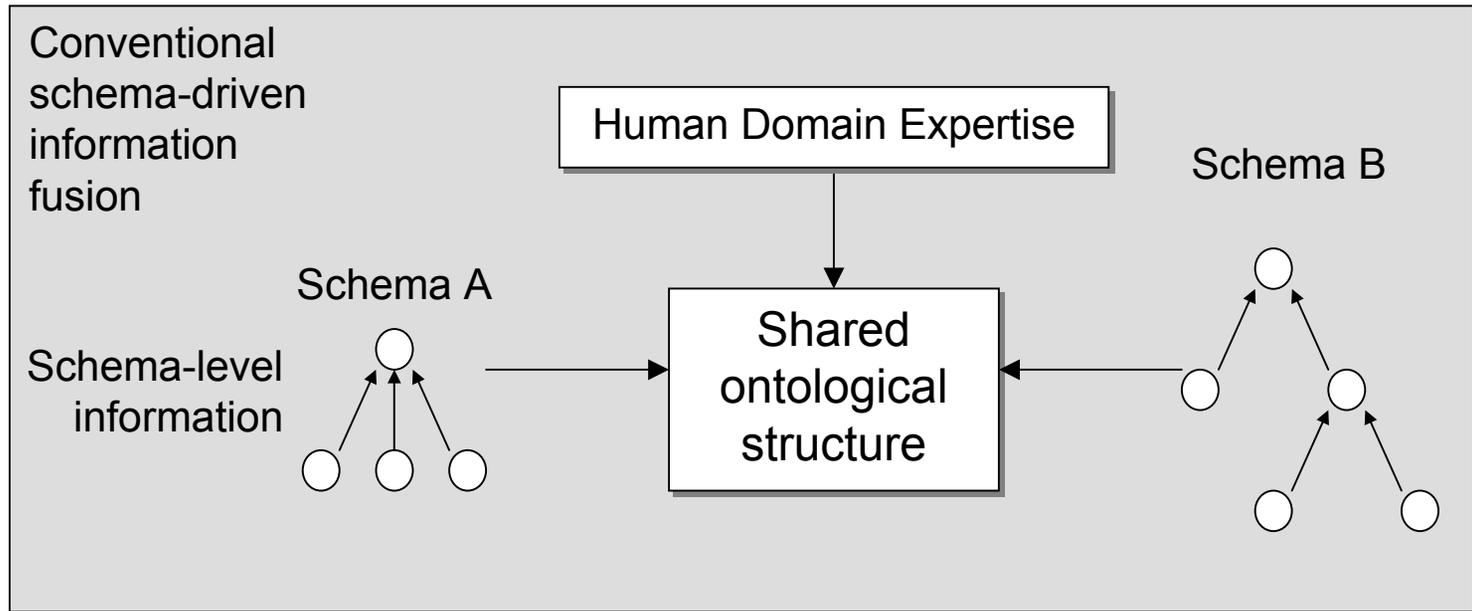
Simple example: schema-level



Schema-level information fusion

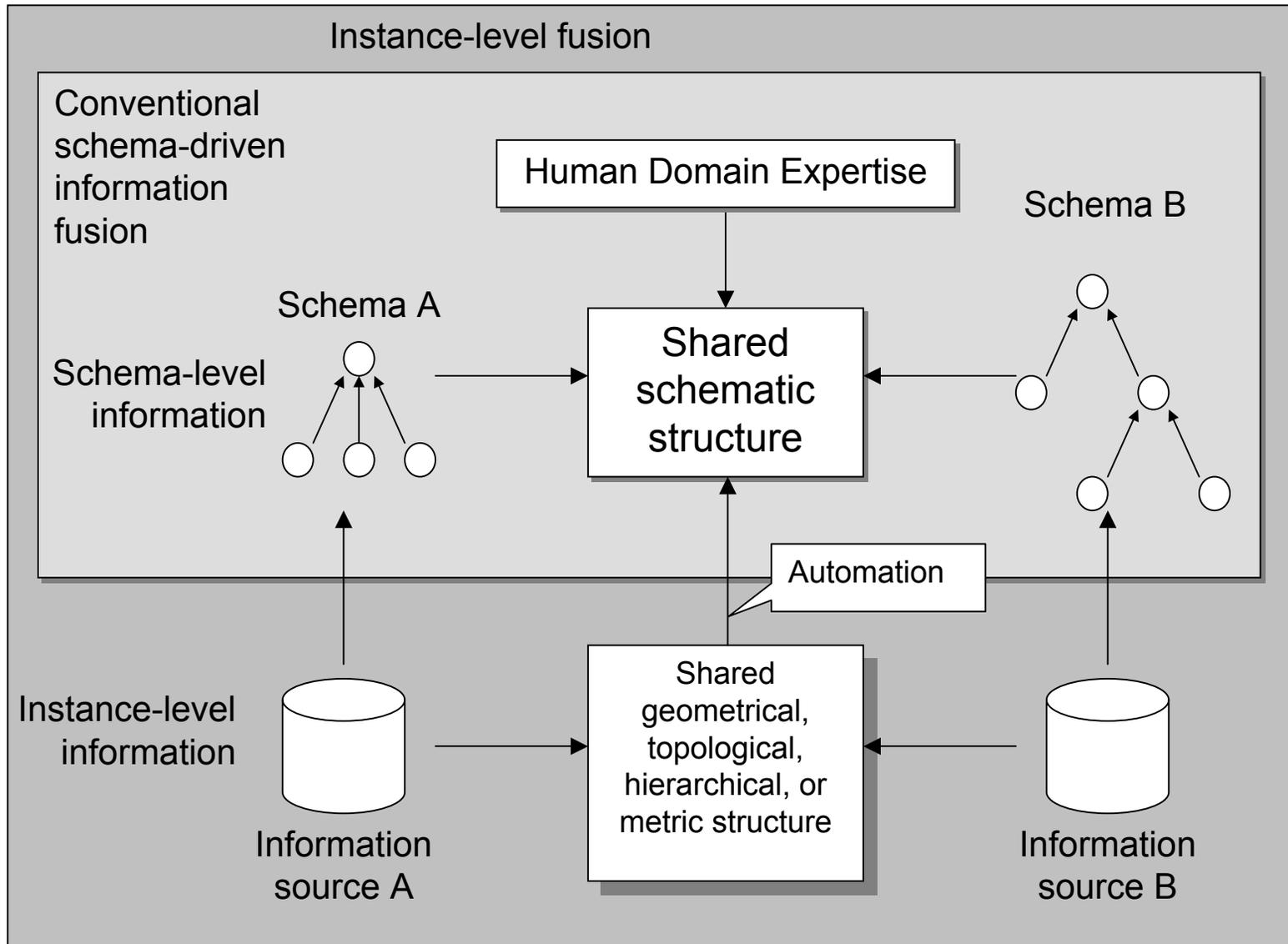
- Databases
 - schema integration (eg Kim & Sea, 1992)
 - mediators (eg Wiederhold, 1992)
- Interoperability (eg Sheth, 1999)
- Semantic web (eg Berners-Lee et al., 2001)
- Data warehousing and IR (eg Widom, 1995)
- Ontology-based IS (Fonseca et al, 2002)
- Knowledge representation (Calvanese et al., 1998)

Schema-level information fusion



- ❑ Concentrating solely on the schema ignores
 - a potentially valuable source of examples
 - the mismatch between how information is defined and structured and how it is actually used
- ❑ High levels of human domain expertise are normally needed to fuse information

Instance-based information fusion



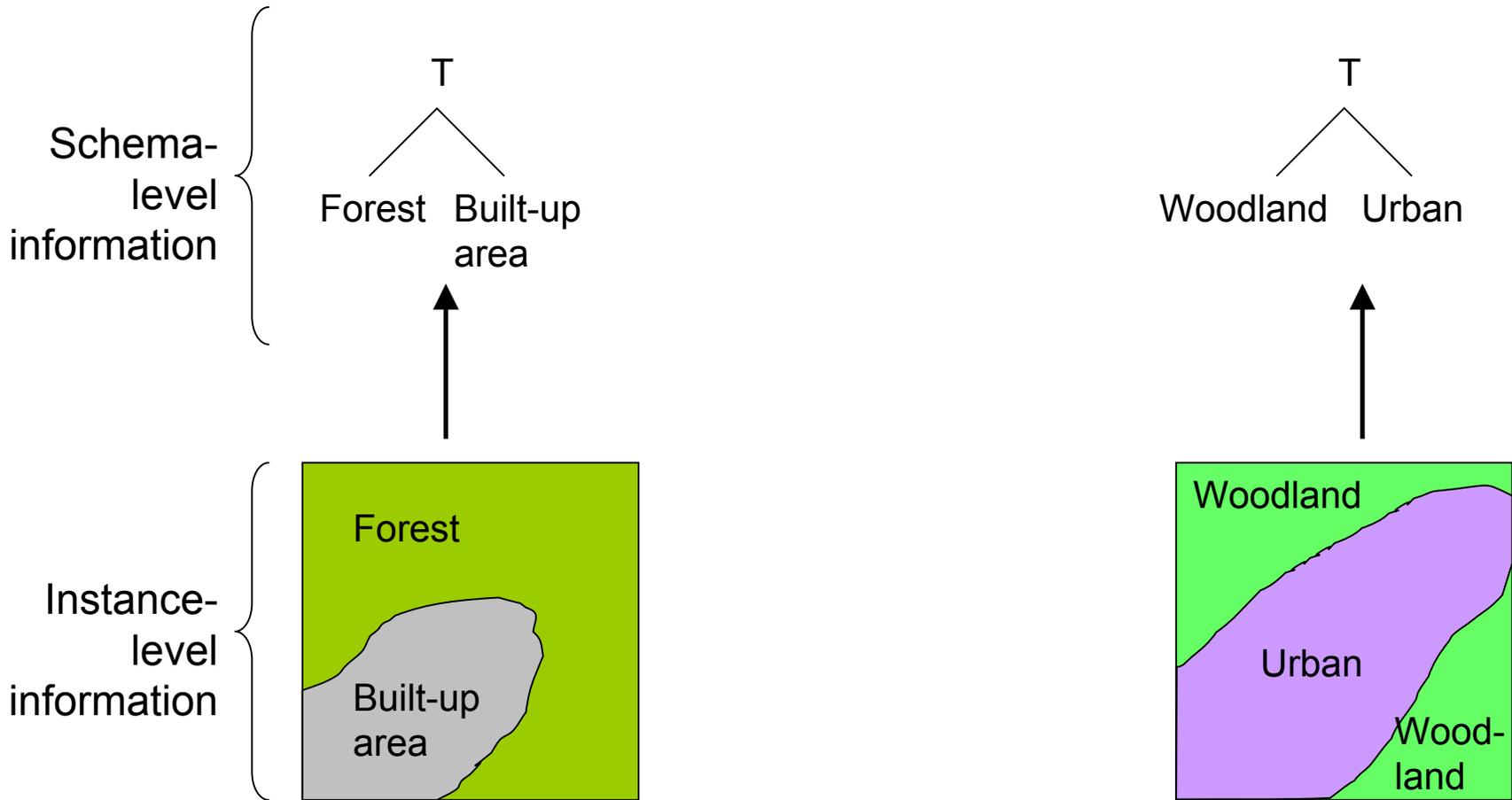
Instance-level information fusion

- Probabilistic approaches in semantic web (Doan et al, 2002)
- Neural nets in databases (Li and Clifton, 2000)
- Machine learning in databases (Berlin and Motro, 2001)

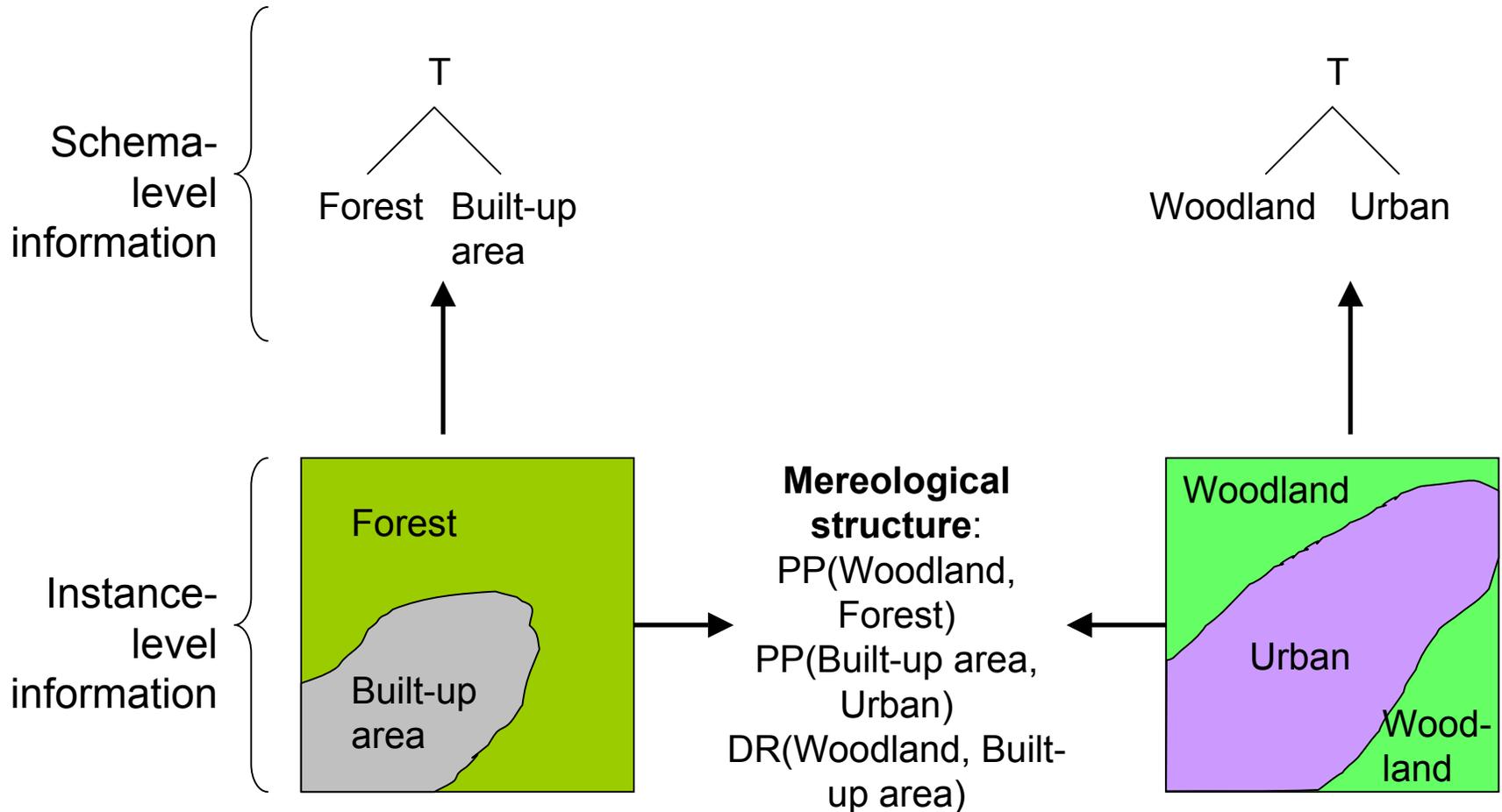
Minimal reasoning system

- Assume we have “good” classifications that induce a partition on observable phenomena
- Devise rule to relating instance-level structure to schema-level structure, e.g.
 - If the extension of a category x is a proper part of the extension of another category y then category x is subsumed by category y

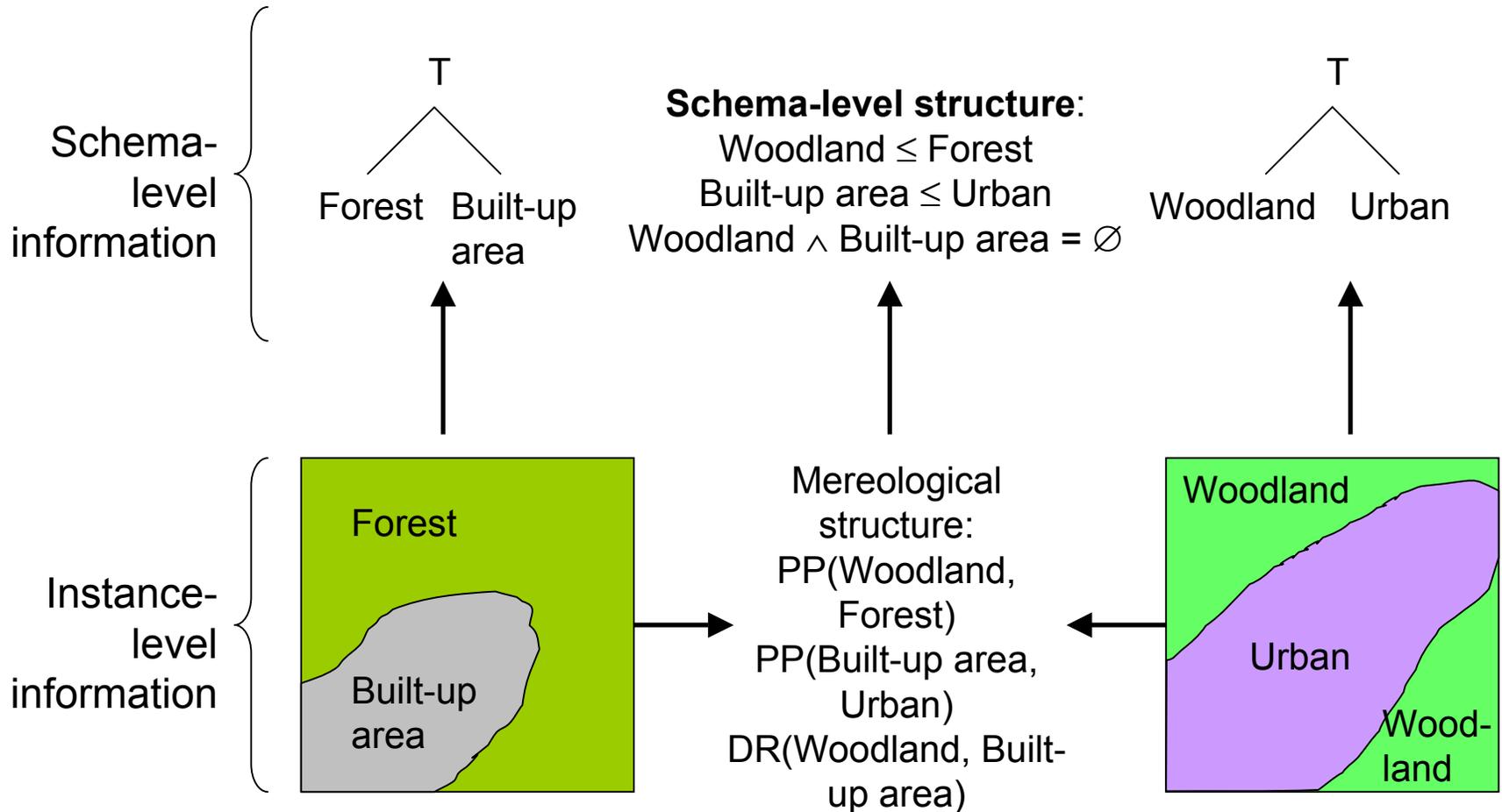
Simple example: instance-level



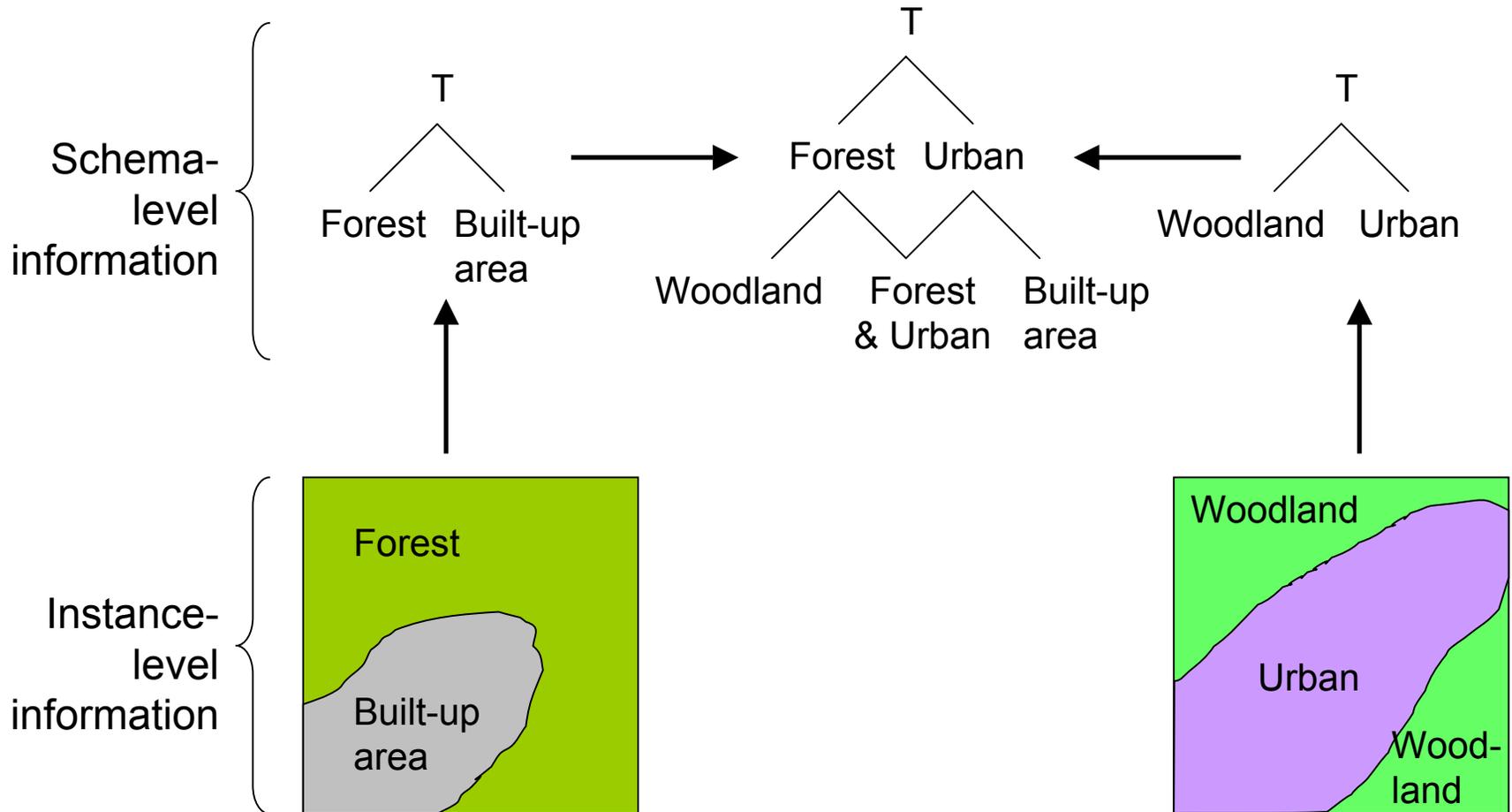
Simple example: instance-level



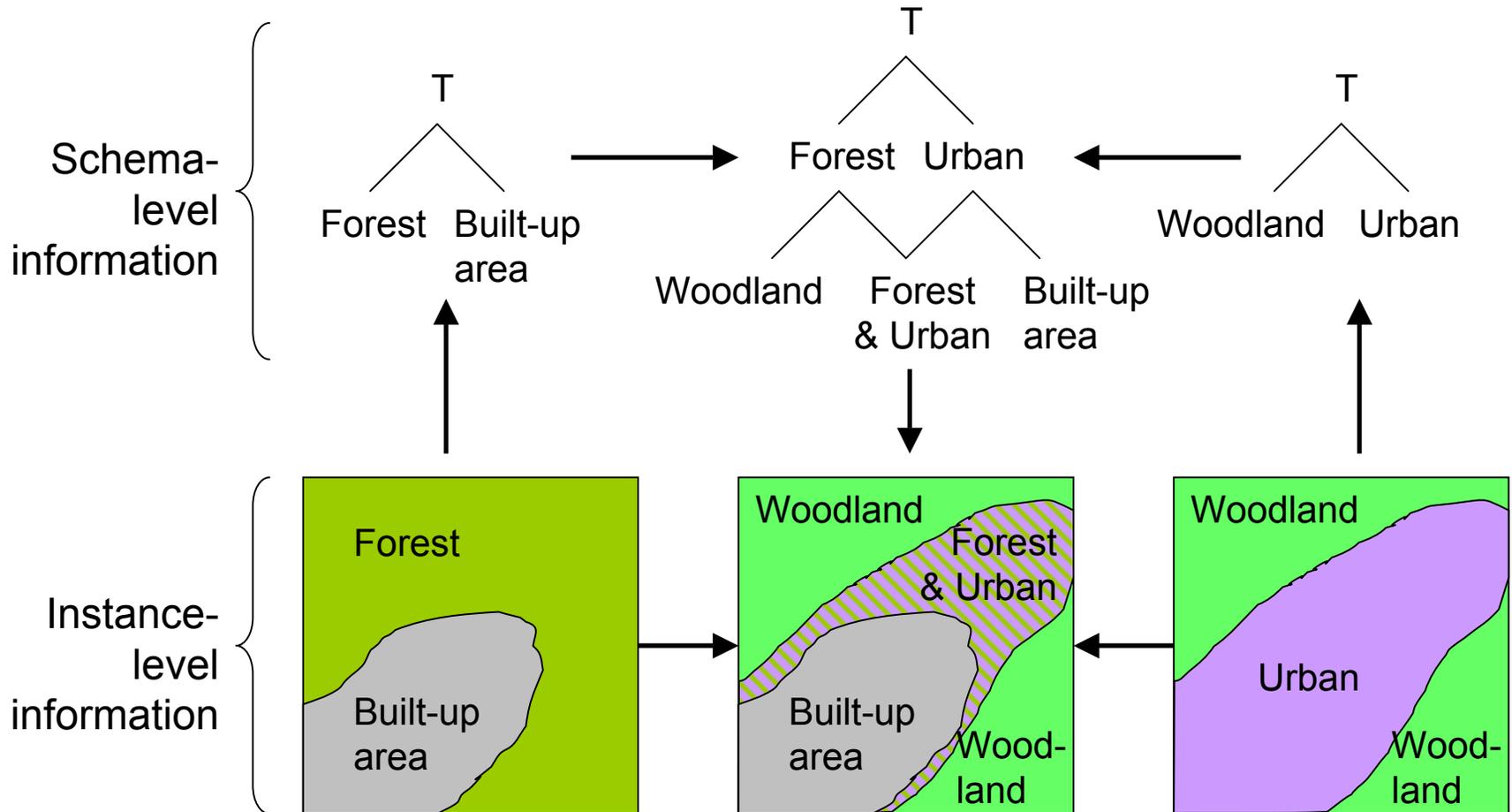
Simple example: instance-level



Simple example: instance-level



Simple example: instance-level



Construction

- Assume $T1$ and $T2$ are meet semilattices.
- Given two functions $g1: T1 \rightarrow R$, $g2: T2 \rightarrow R$, the first step is to build the fused hierarchy.
- Take $T1 \times T2$ as the set of ordered pairs, where:
 - $(s1, s2) \wedge (t1, t2) = (s1 \wedge s2, t1 \wedge t2)$
 - (we also have to worry a little about \perp)
- How does this instantiate?
 - $g(t1, t2) = g_{t1} \wedge g_{t2}$ (in the lattice of partitions)
- Factor $T1 \times T2$ by a set of constraints
 - e.g. $t =< t'$ iff $g_t =< g_{t'}$

Uncertainty

- Real data sets rarely match up so nicely, as a result of uncertainty
 - Vagueness
 - Imprecision
 - Inaccuracy

Logics of fusion under uncertainty

fuse	1	?	0
1	1	?	?
?	?	?	?
0	?	?	0

Vague region
no errors
conservative

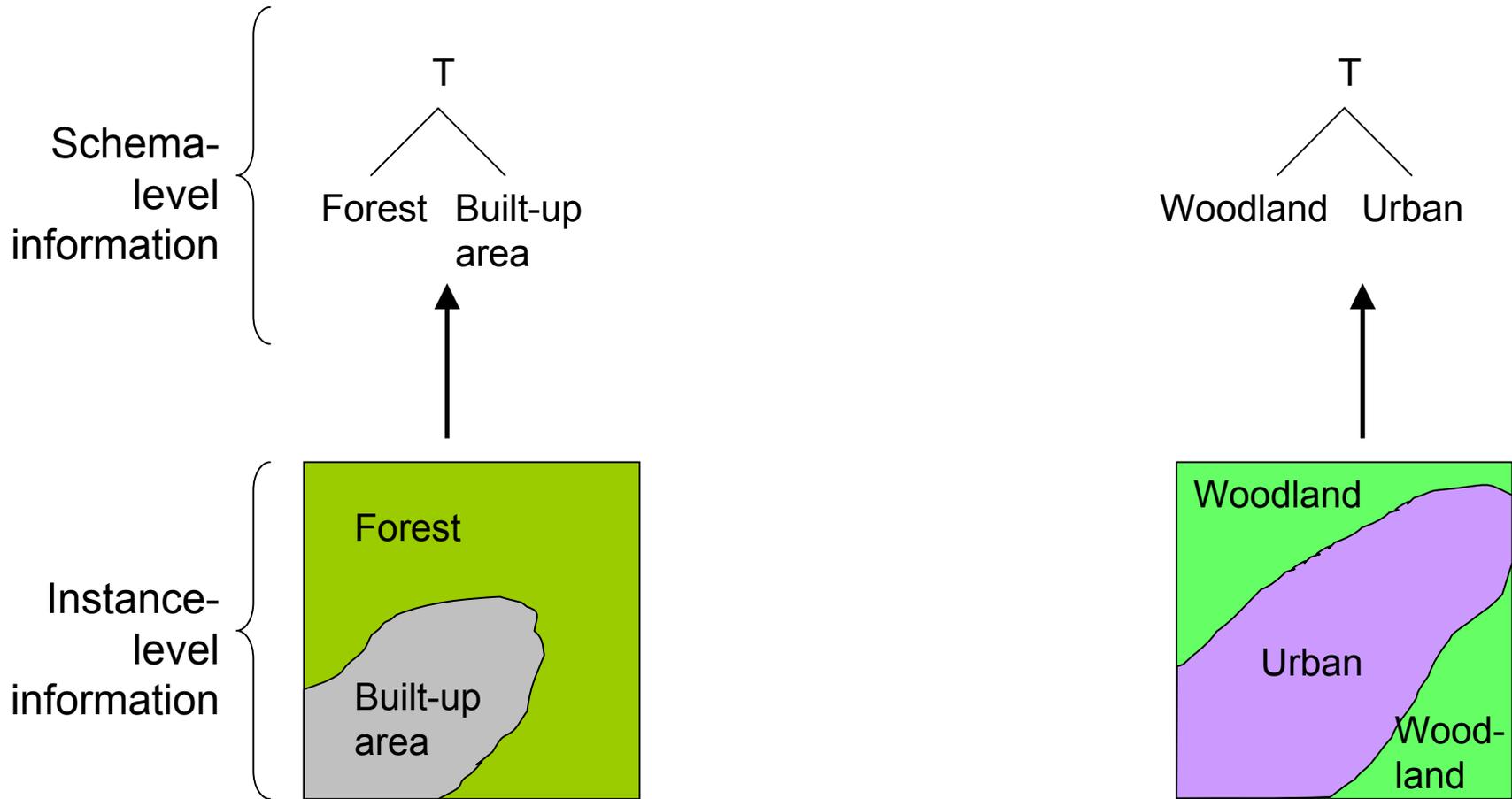
fuse	1	?	0
1	1	1	X
?	?	1	0
0	X	0	0

Crisp region viewed with
imprecision
no errors
weighted towards definite

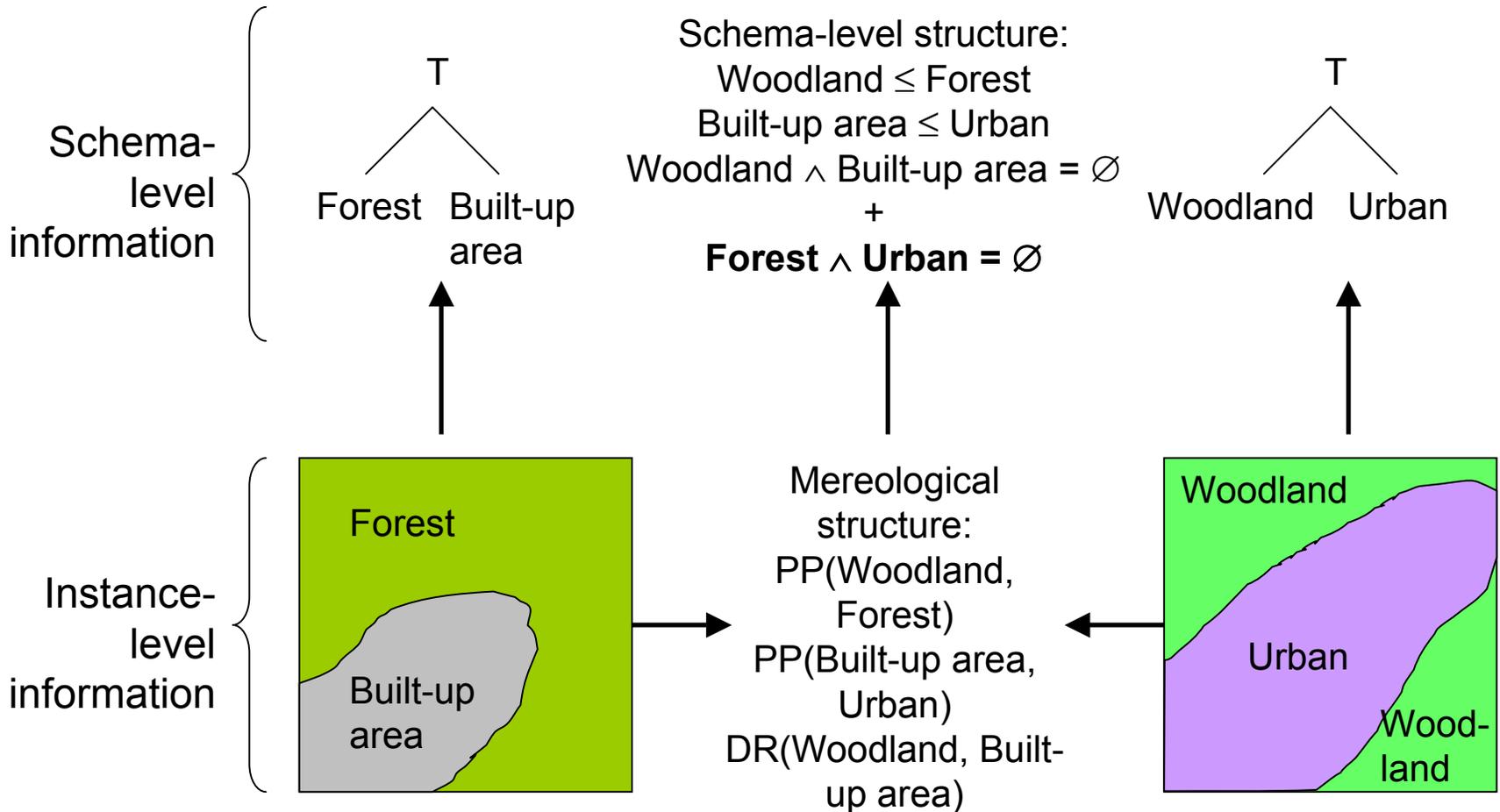
Inaccurate schema

- We want to incorporate schema-level knowledge in fusion process
- This knowledge may conflict with instance-level information

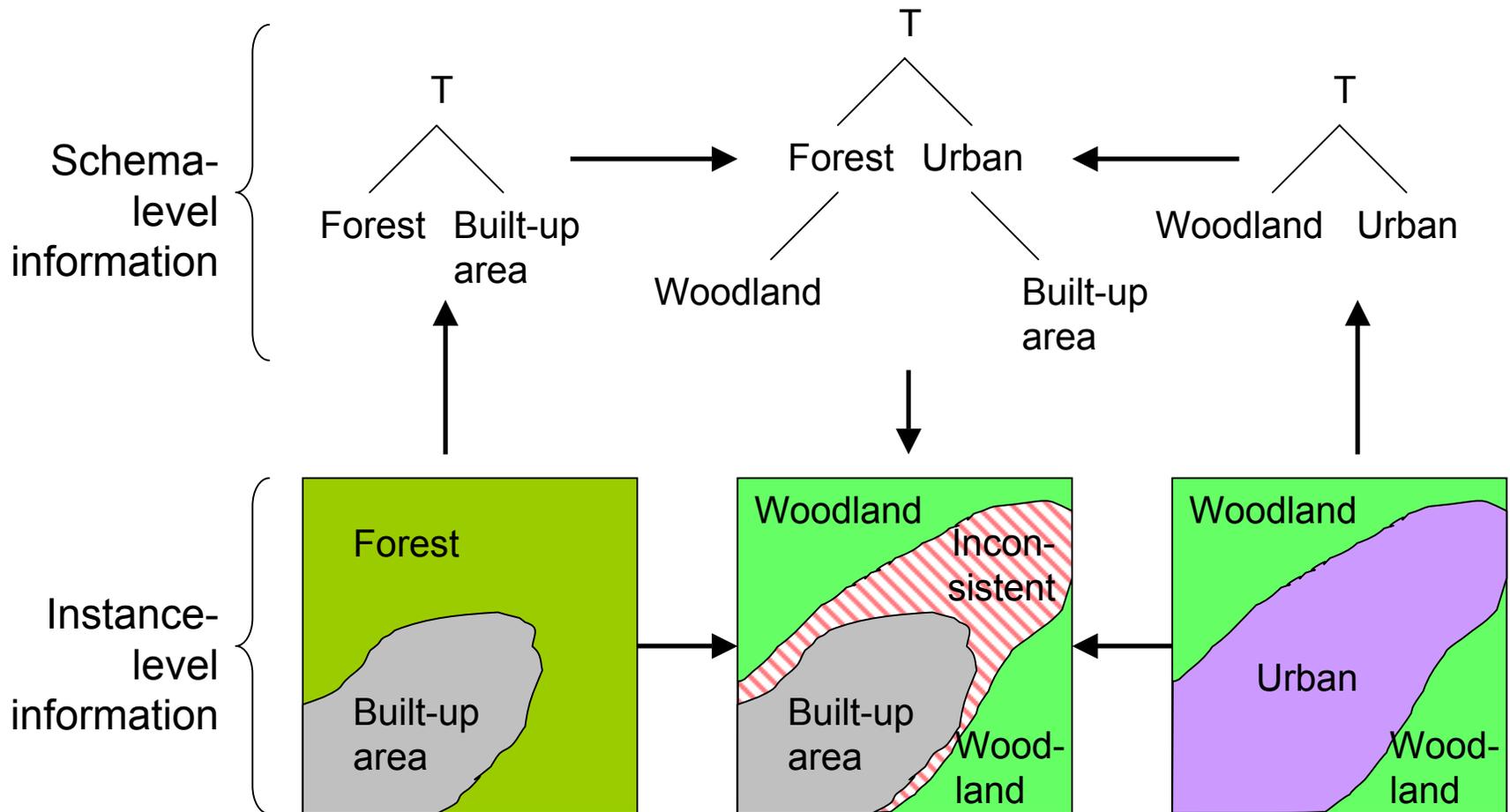
Inaccurate schema



Inaccurate schema

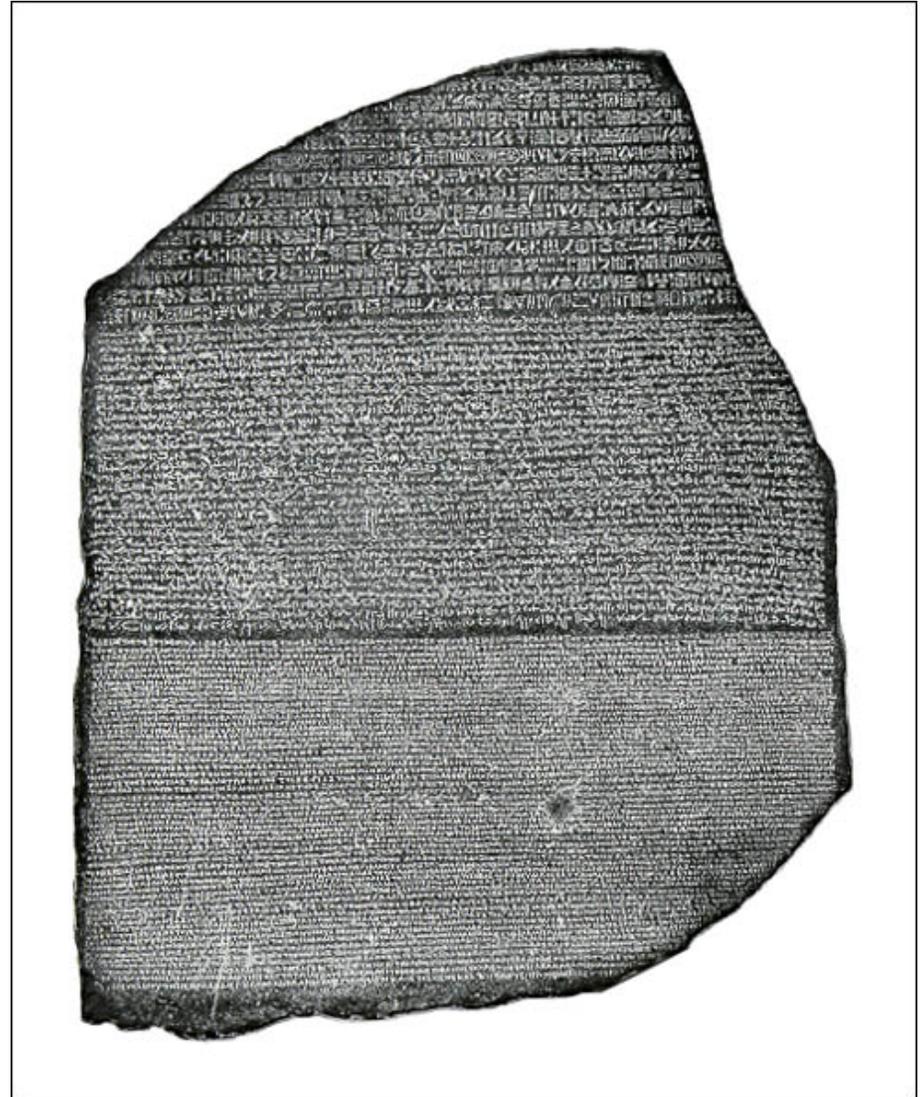


Inaccurate schema

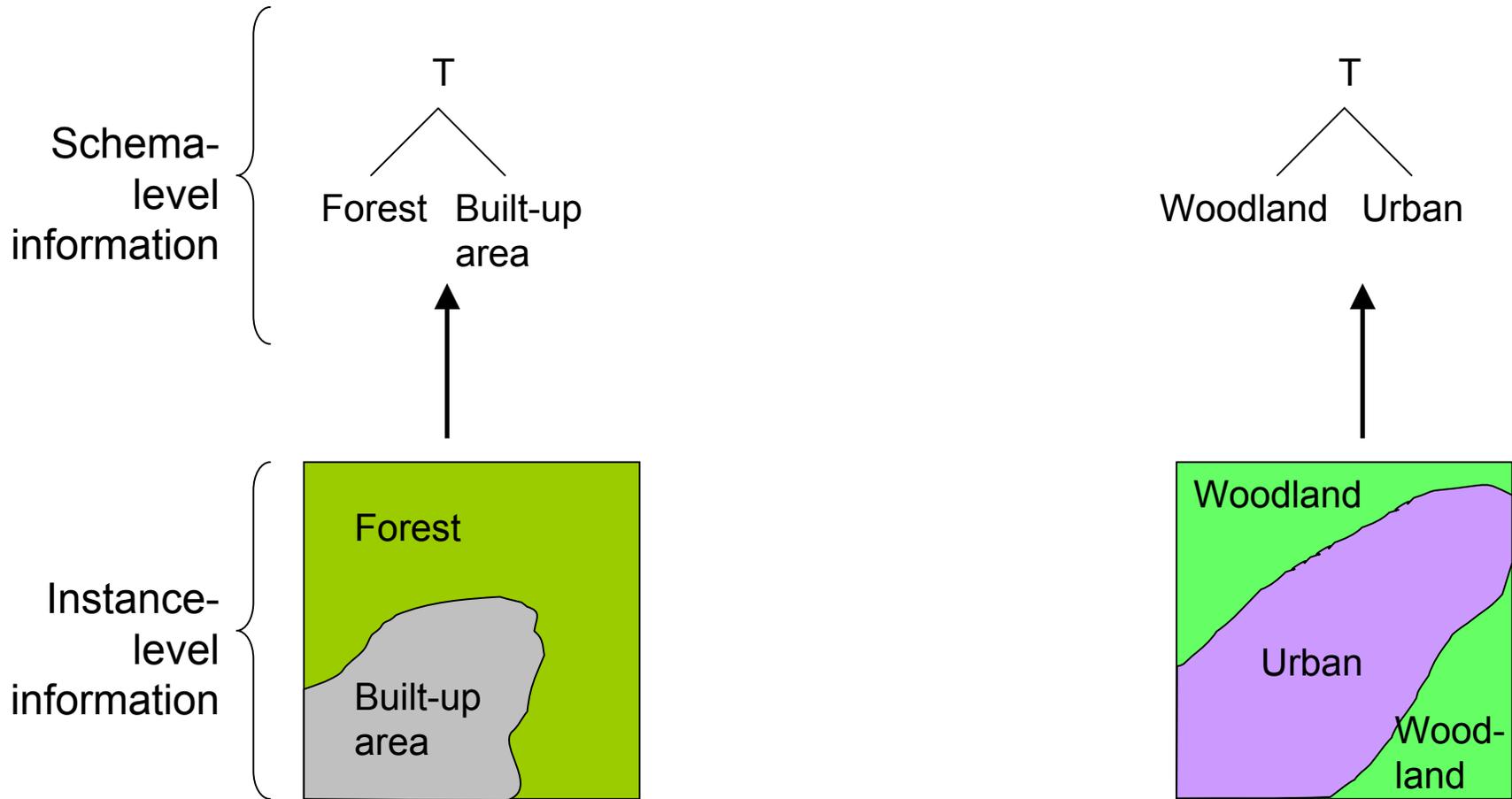


Inaccurate instances

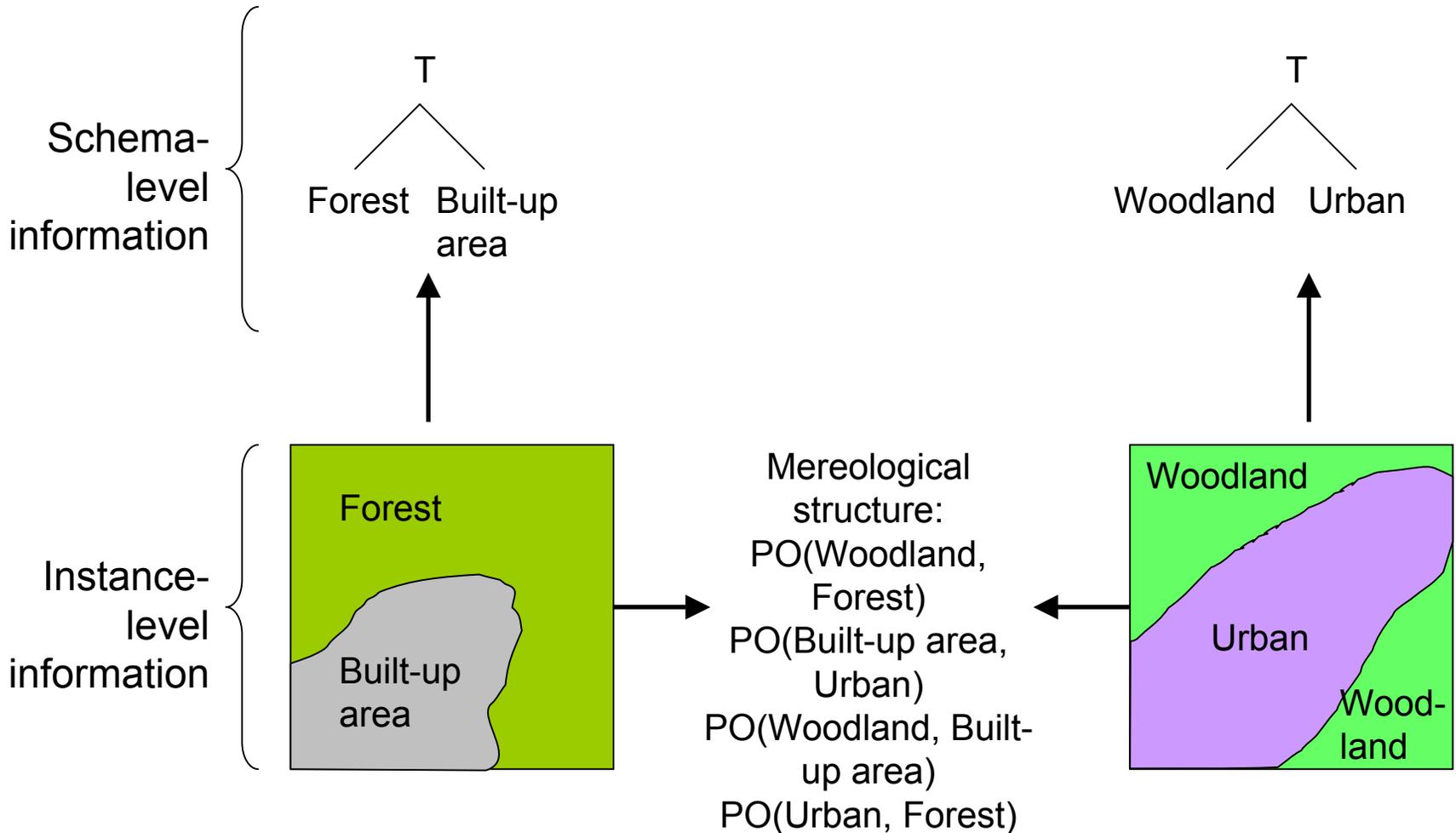
- ❑ What if one of the three texts had a typo?
- ❑ With example data sets, we don't normally rely solely on a single instance to draw conclusions
- ❑ So if *most* instances agree, perhaps we could ignore errant instances like a typo



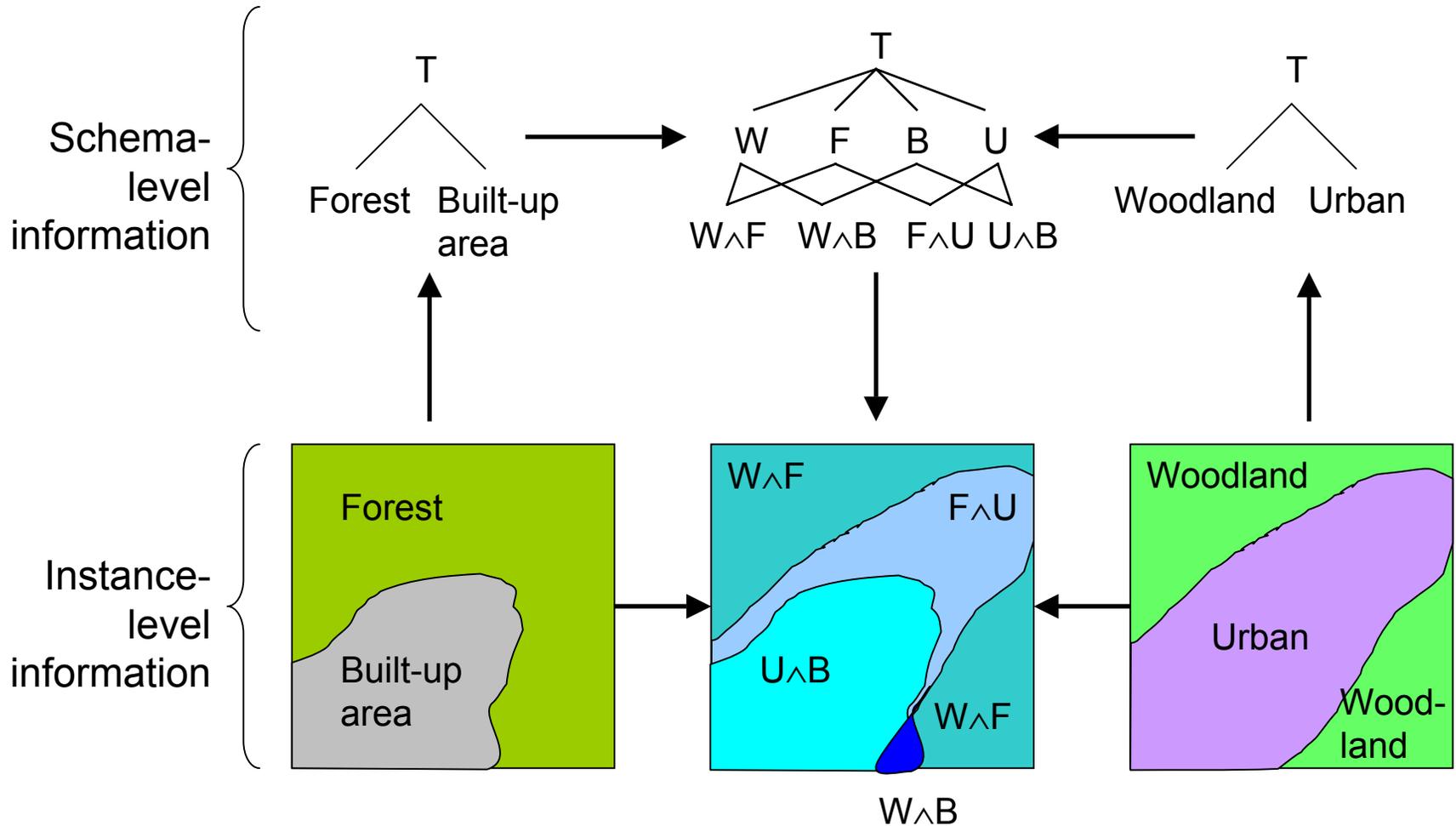
Inaccurate instances



Inaccurate instances

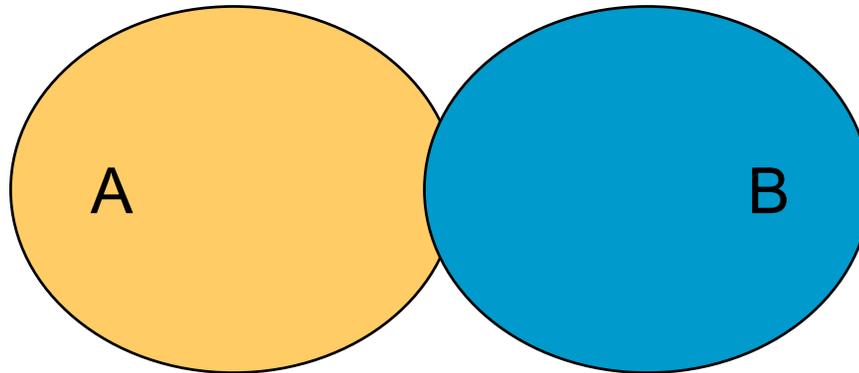


Inaccurate instances



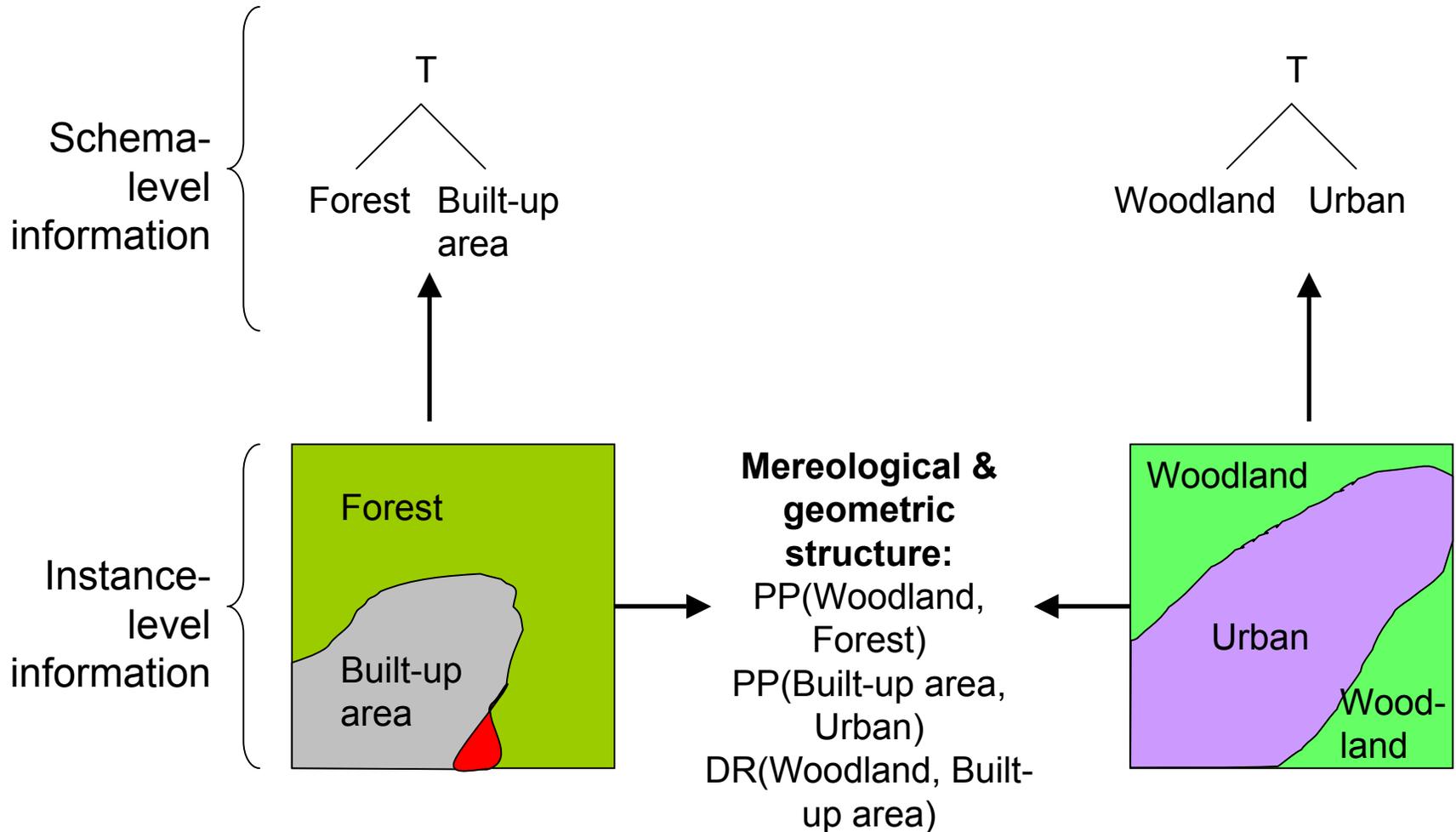
Inaccurate instances

- Assume that extensions that overlap to a relatively small extent are due to inaccuracies
- In our example, include information about geometric structure

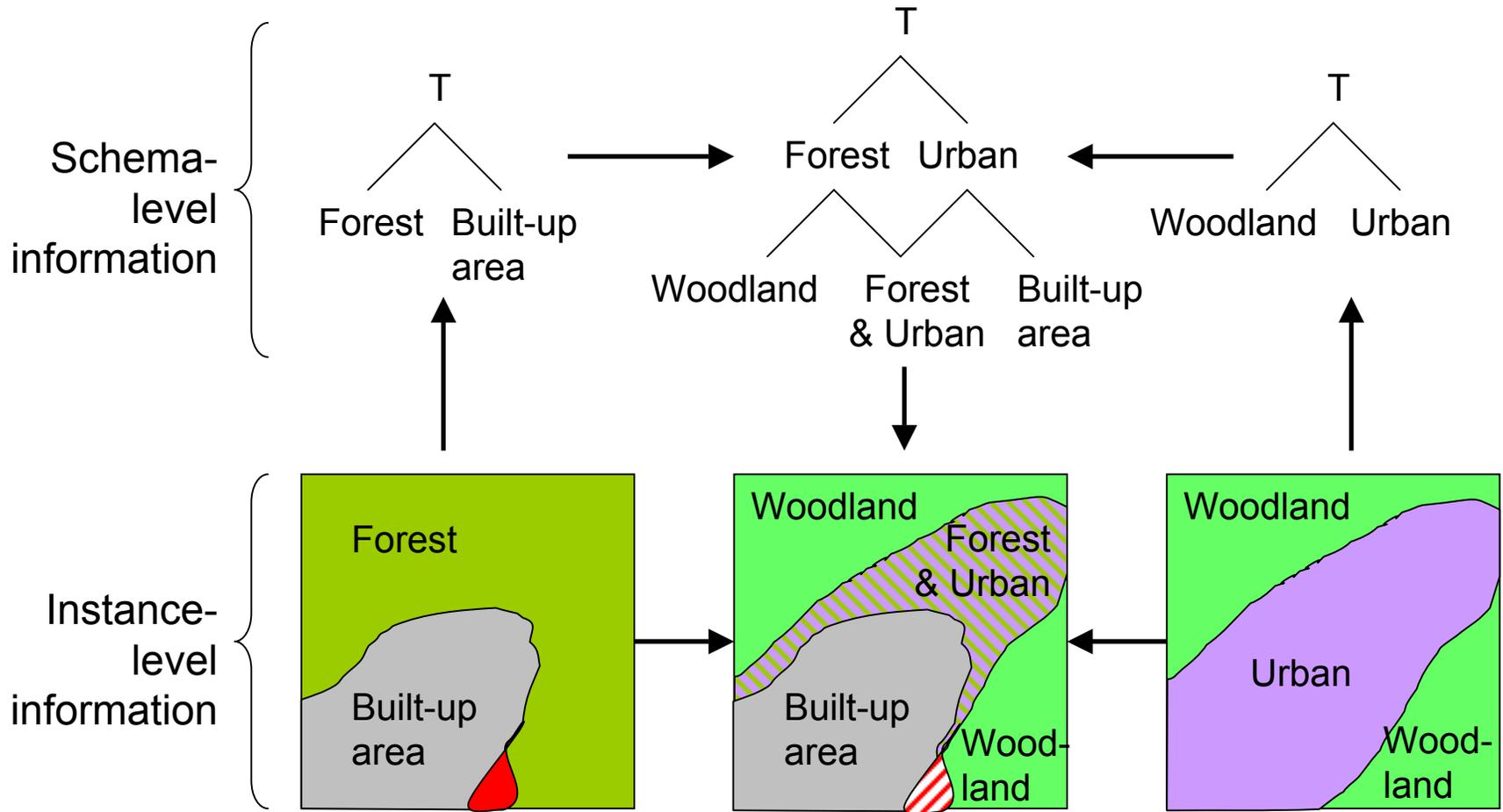


- Eg, if $\text{Area}(A \text{ overlaps } B) / \text{Area}(A) < \delta$
& $\text{Area}(A \text{ overlaps } B) / \text{Area}(B) < \delta$
ignore overlap during inference process

Inaccurate instances

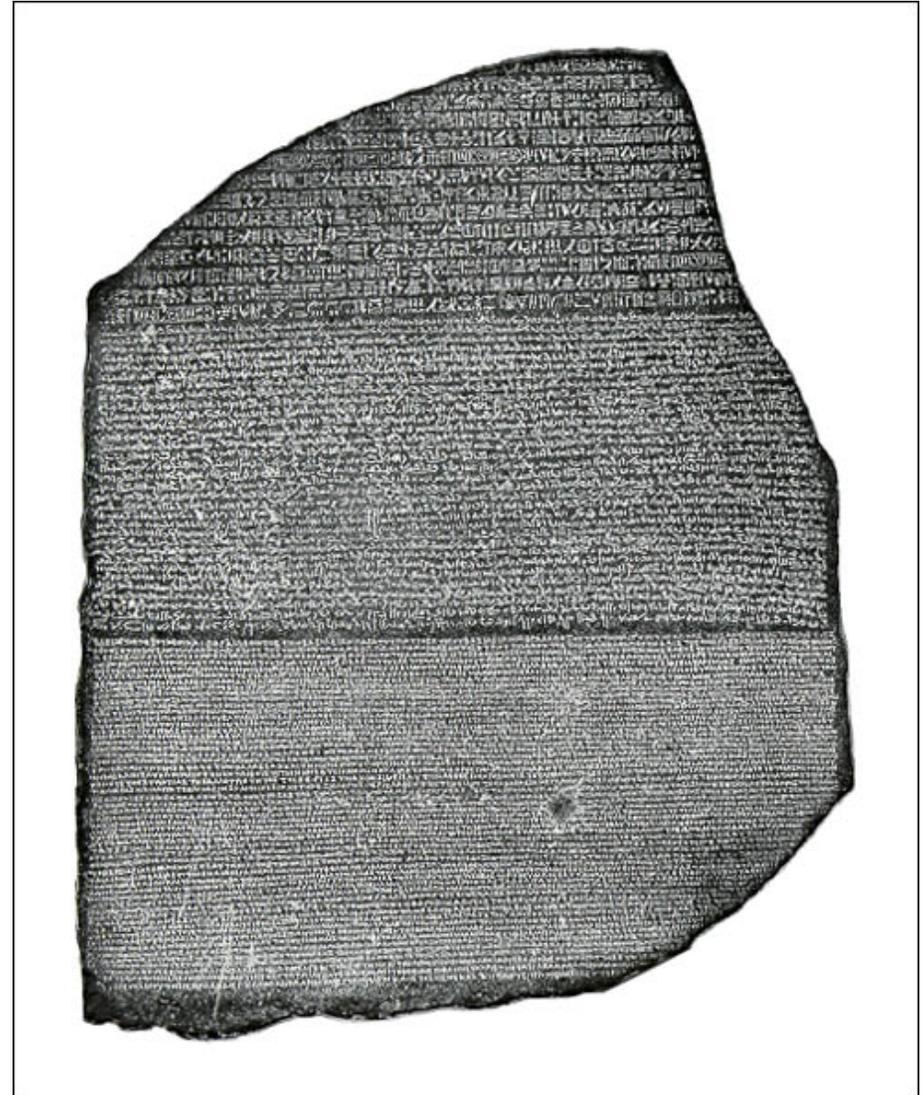


Inaccurate instances



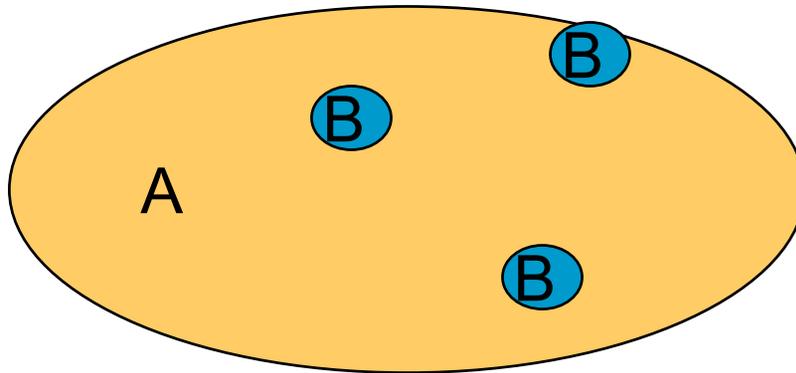
Imprecise instances

- ❑ What if one of the three texts had more detail than the others?
- ❑ If we have qualitative knowledge about the relative granularities of text, perhaps we can allow for more detail, and not try to use *everything* in the more detailed text for inferring schema structure



Imprecise instances

- Assume that extensions that overlap to a widely different extent are due to different levels of precision in information sources

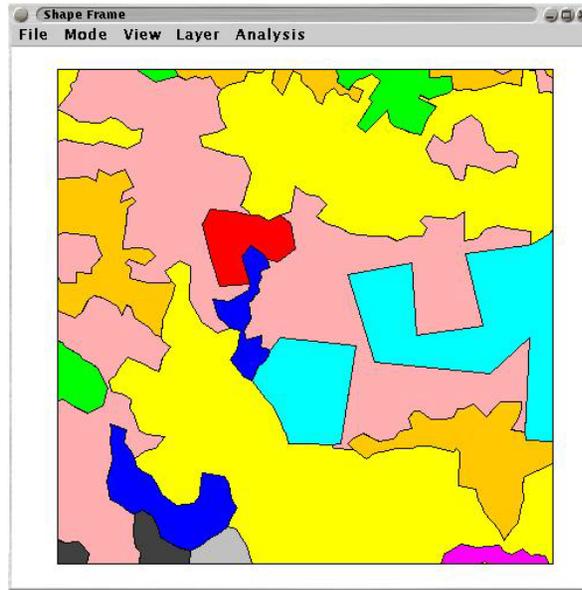


Eg, B="Hamlet" in fine grained data set, A="Woodland" in coarse grained data set

- Eg, if $\text{Area}(A \text{ overlaps } B) / \text{Area}(A) < \delta$ & $\text{Area}(A \text{ overlaps } B) / \text{Area}(B) > 1 - \delta$ ignore overlap

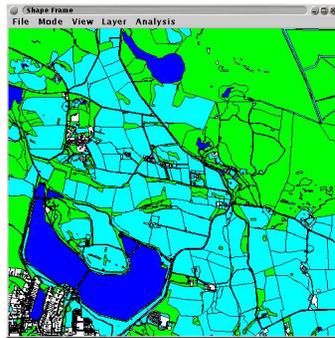
Prototype implementation

Land Cover Data (eg DNF&CORINE)

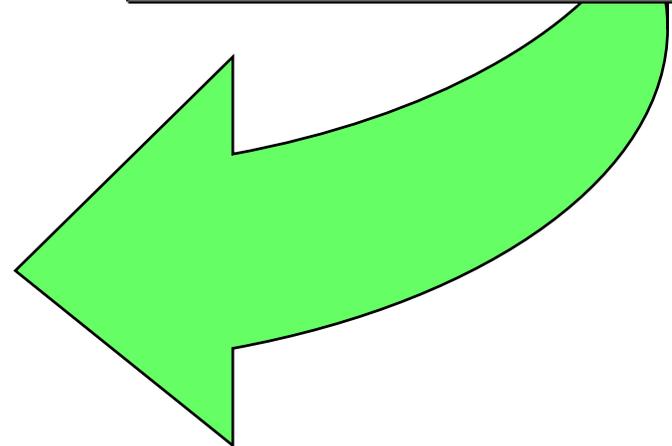
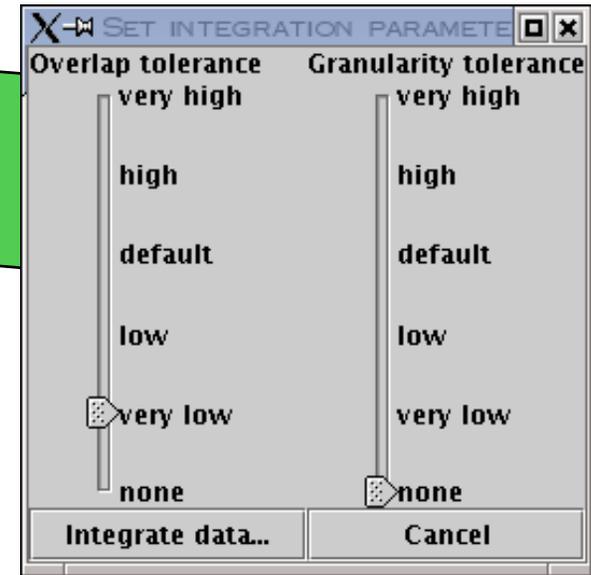
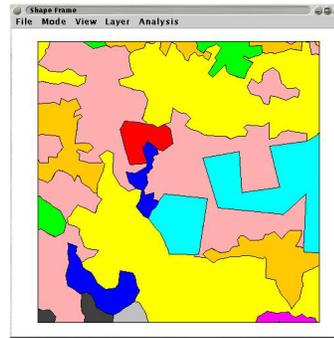


Prototype implementation

Land Cover Data (eg DNF&CORINE)

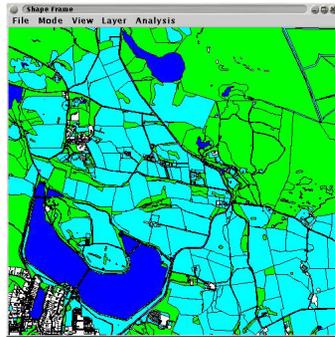


+

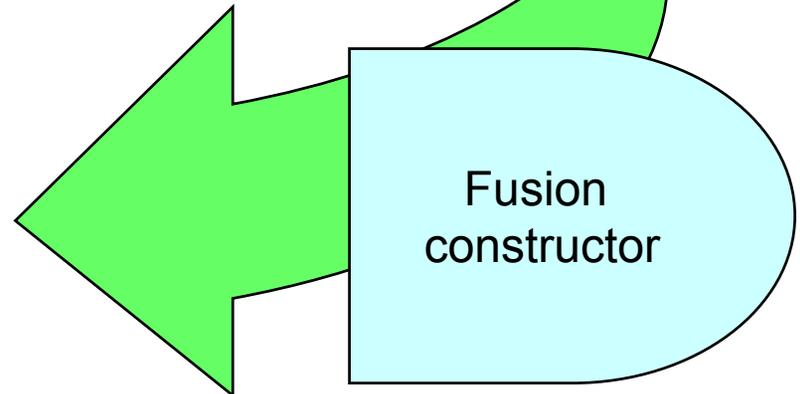
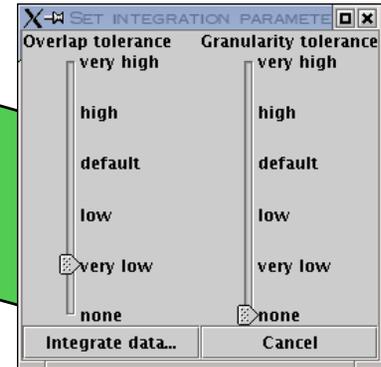
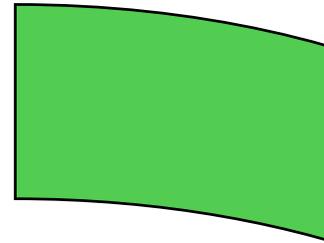
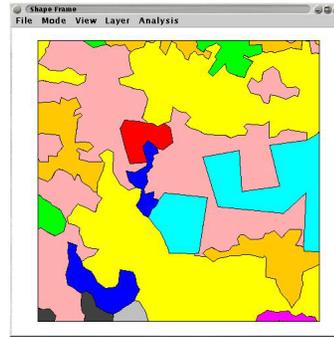


Prototype implementation

Land Cover Data (eg DNF&CORINE)



+

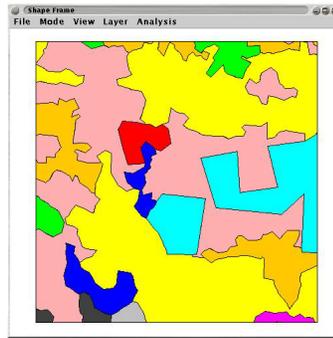


Prototype implementation

Land Cover Data (eg DNF&CORINE)



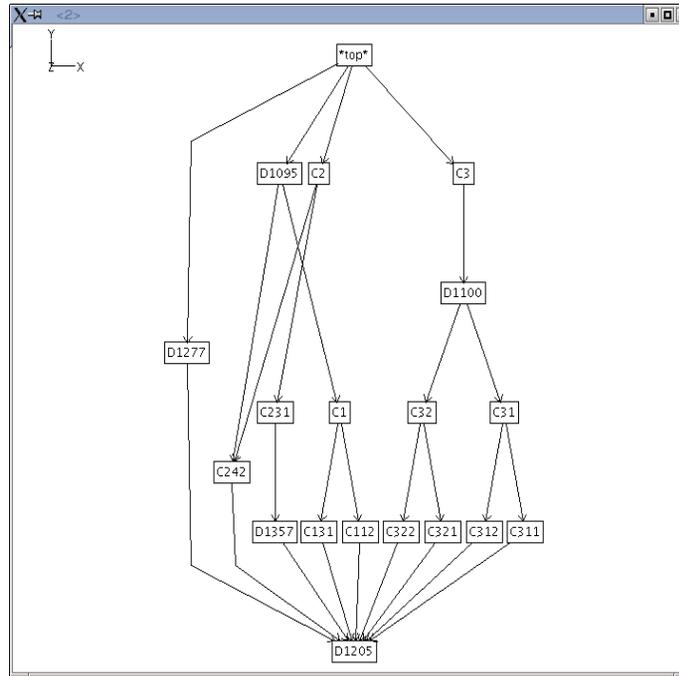
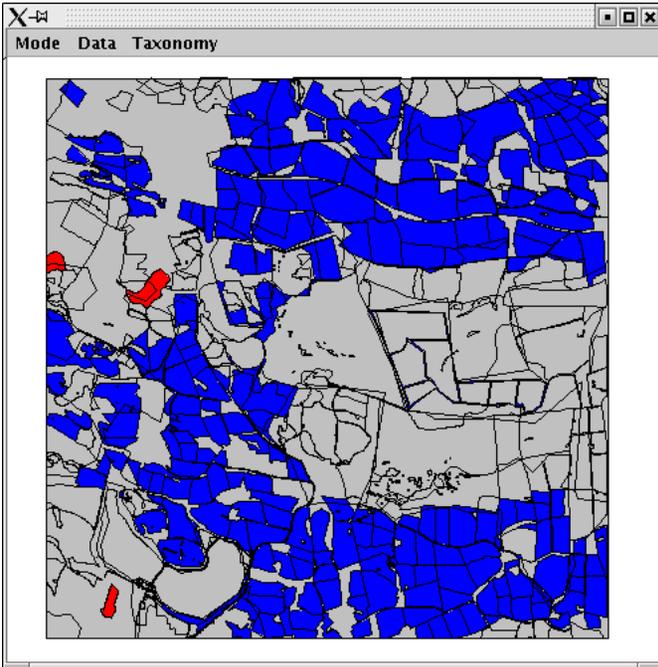
+



SET INTEGRATION PARAMETER

Overlap tolerance	Granularity tolerance
very high	very high
high	high
default	default
low	low
very low	very low
none	none

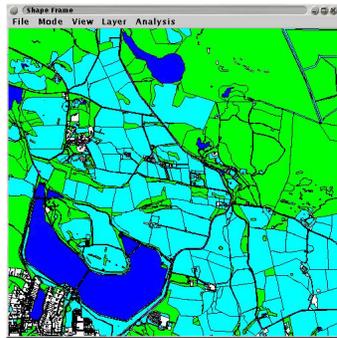
Integrate data... Cancel



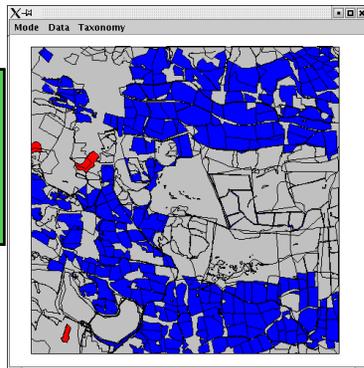
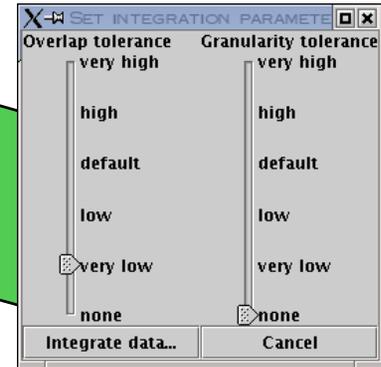
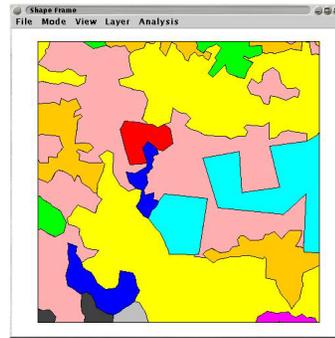
Fusion constructor

Prototype implementation

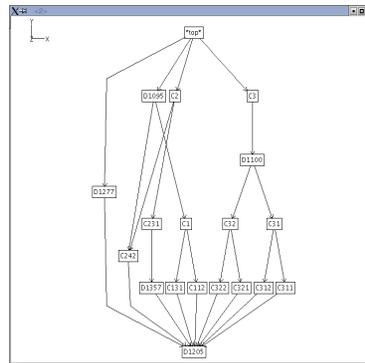
Land Cover Data (eg DNF&CORINE)



+



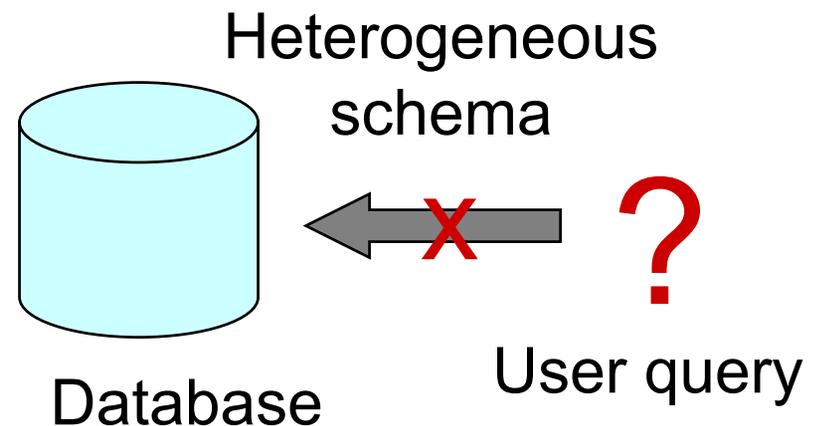
+



Fusion constructor

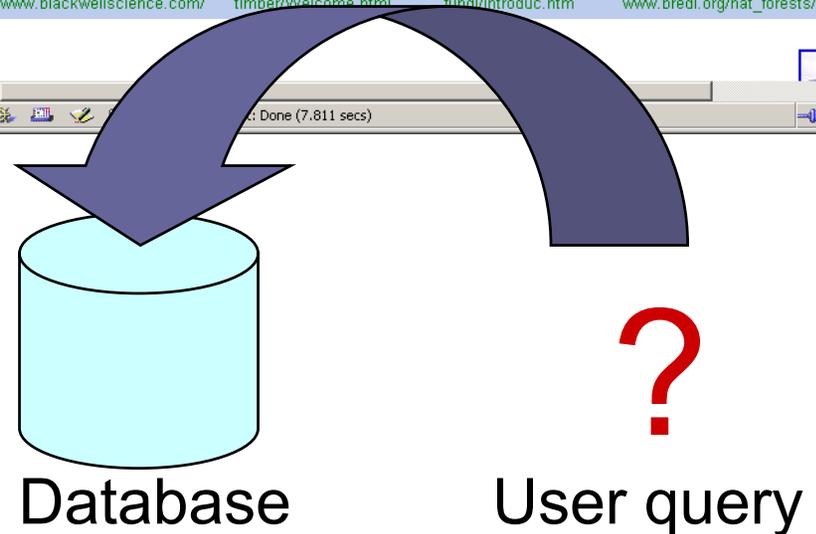
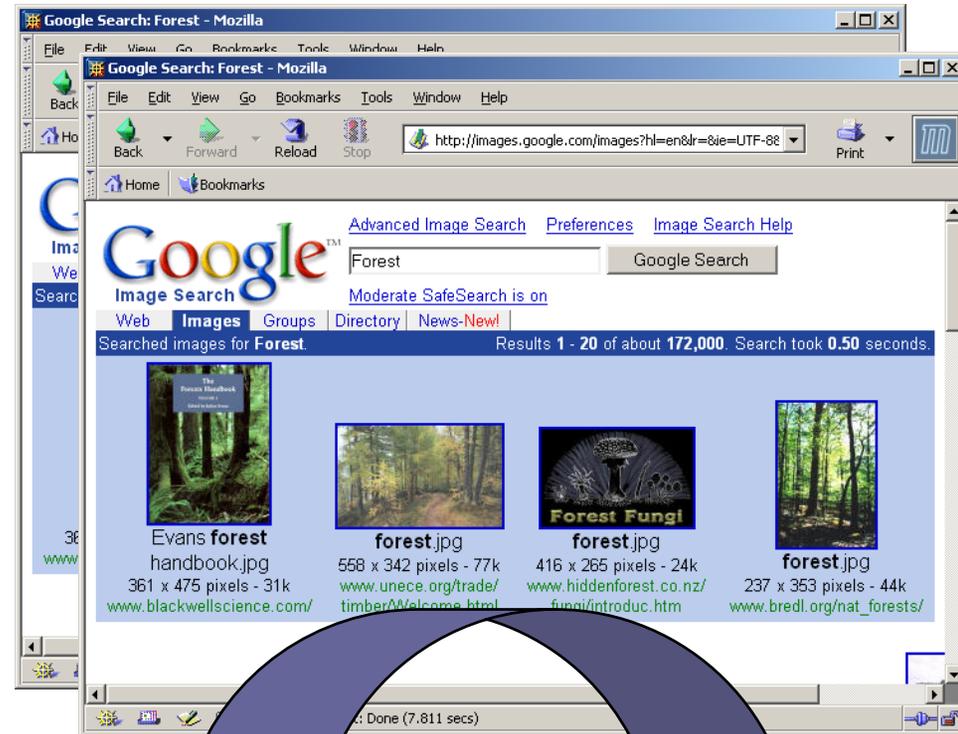
Other structures: instance-level

- ❑ Other instance-level structures: hierarchical and similarity
- ❑ E.g. “Internet as Thesaurus”
- ❑ Use hierarchical structure of web pages
- ❑ Use similarity structure of text
- ❑ Use similarity structure of images



Other structures: instance-level

- ❑ Other instance-level structures: hierarchical and similarity
- ❑ E.g. “Internet as Thesaurus”
- ❑ Use hierarchical structure of web pages
- ❑ Use similarity structure of text
- ❑ Use similarity structure of images



Conclusions

- ❑ Trying to take advantage of the way that categories are used to inform the arrangement of schemas.
- ❑ Granularity needs special attention.
- ❑ Problems with conflation of mereological and subsumption structures?
- ❑ Preference relations are important, as this is really a problem of belief revision.