

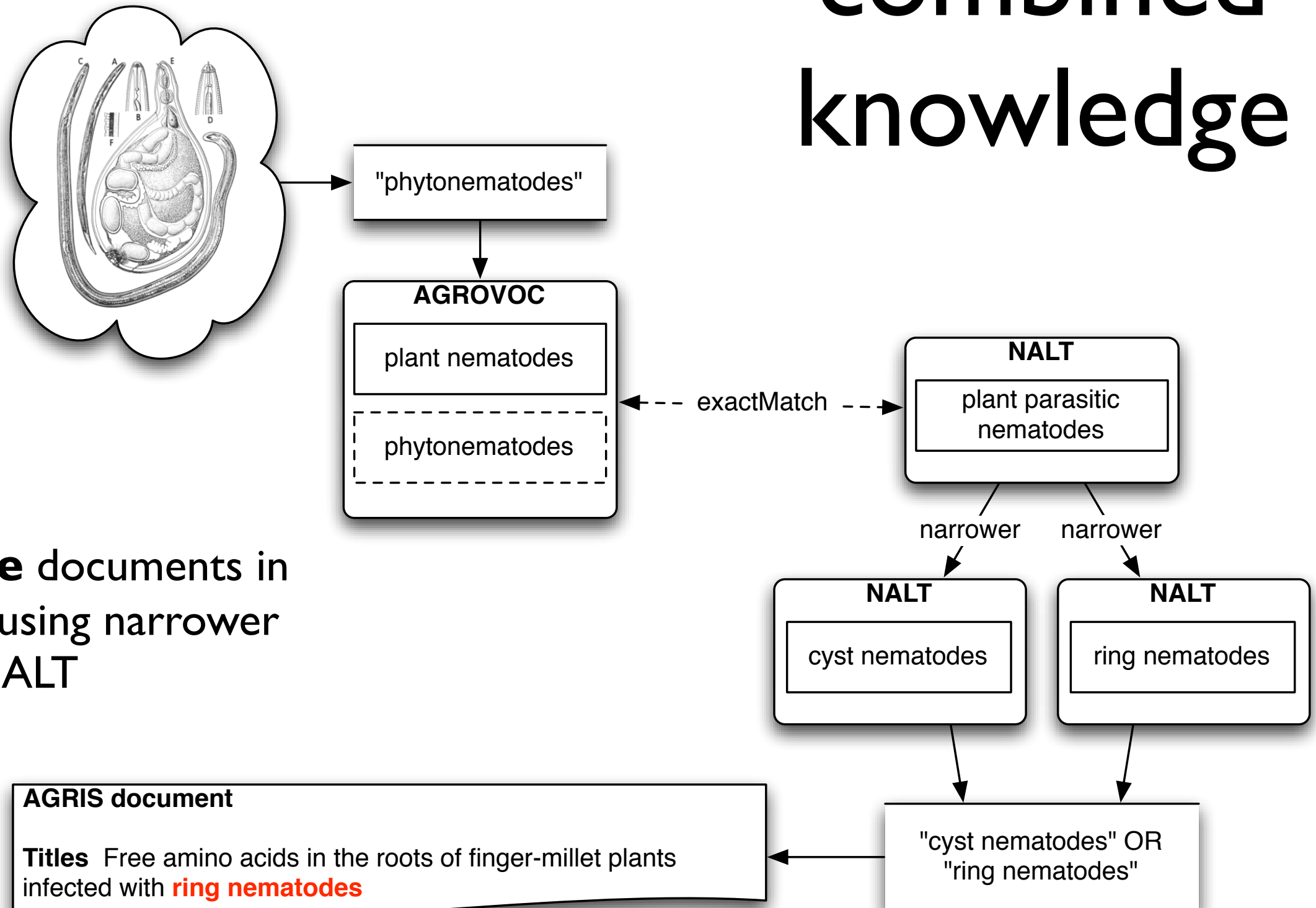
Evaluating Ontology Alignment Techniques

Willem Robert van Hage
VU University Amsterdam

what did I study?

- text mining techniques to find BT (subclass and part-whole) relations in text
 - using respectively NAL, FAO; and FDA, EPA, and WHO data
- various sample-based evaluation techniques
 - end-to-end application evaluation versus stratified sampling
- the quality of current state of the art thesaurus alignment techniques
 - together with NAL, FAO, and EEA
 - at NKOS 2008 Lori Finch talked about our work on comparative evaluation tasks at the OAEI 2006/2007

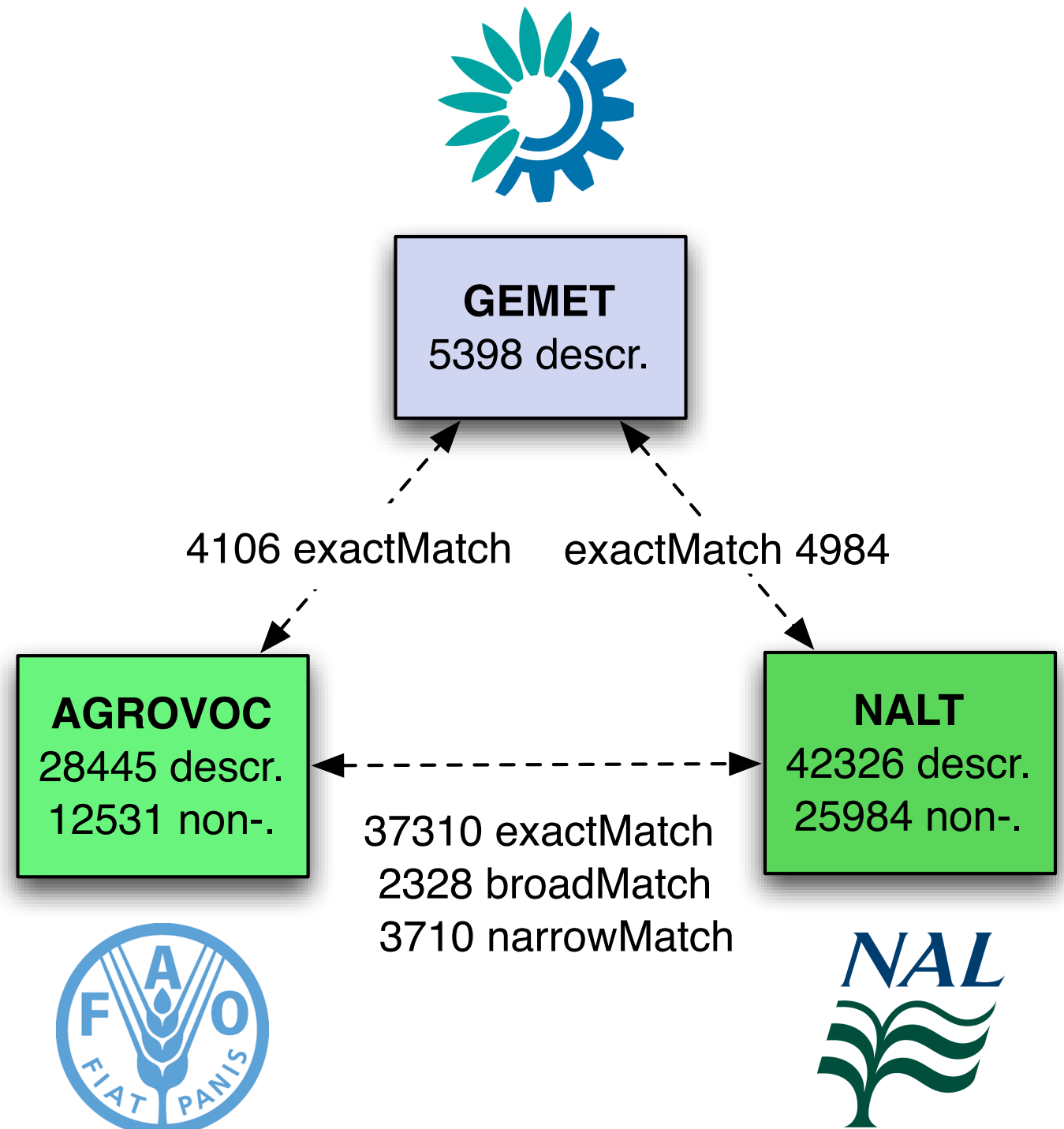
exploit combined knowledge



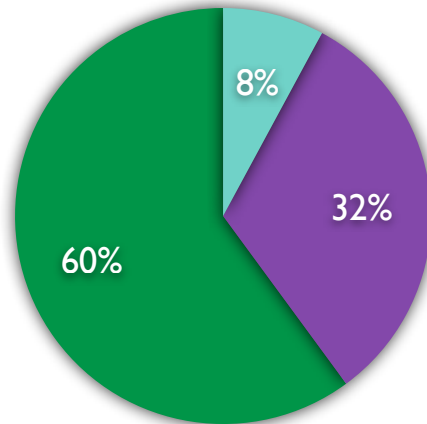
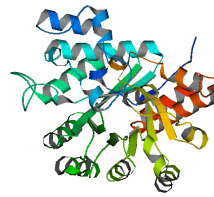
retrieve **more** documents in
FAO's AGRIS using narrower
terms from NALT

some numbers

- OAEI 2007 food & environment tasks (fully automatic)
- mostly but not only skos:exactMatch
- sample evaluation ± 1650 mappings

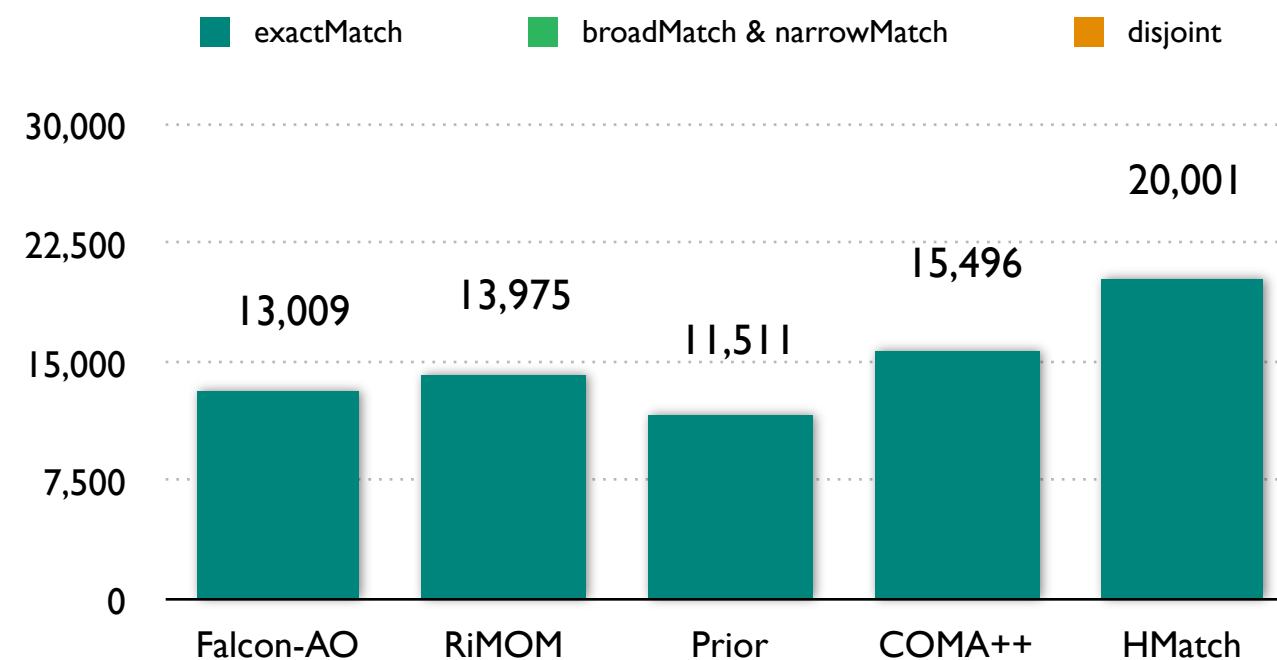
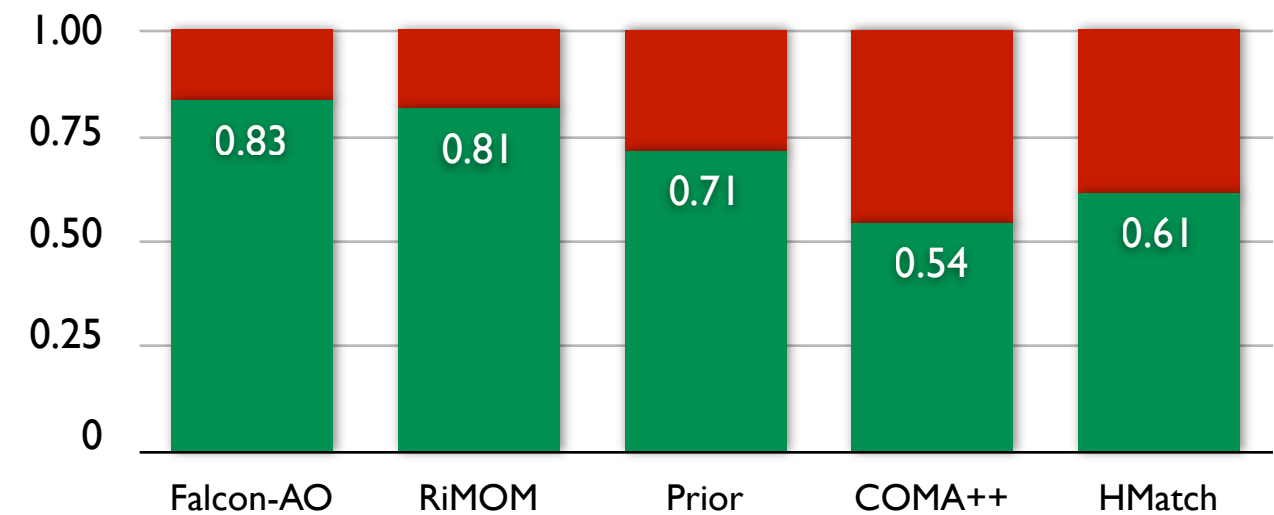


OAEI 2006 food task

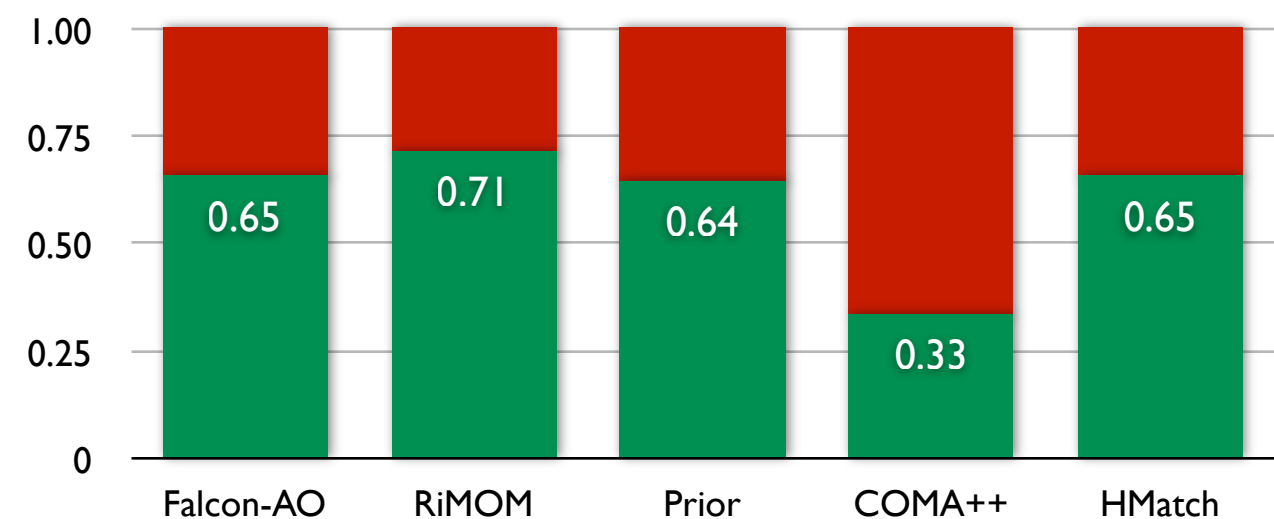


- biological & chemical
- miscellaneous (geographical, legislation, food stuffs, etc.)
- taxonomical

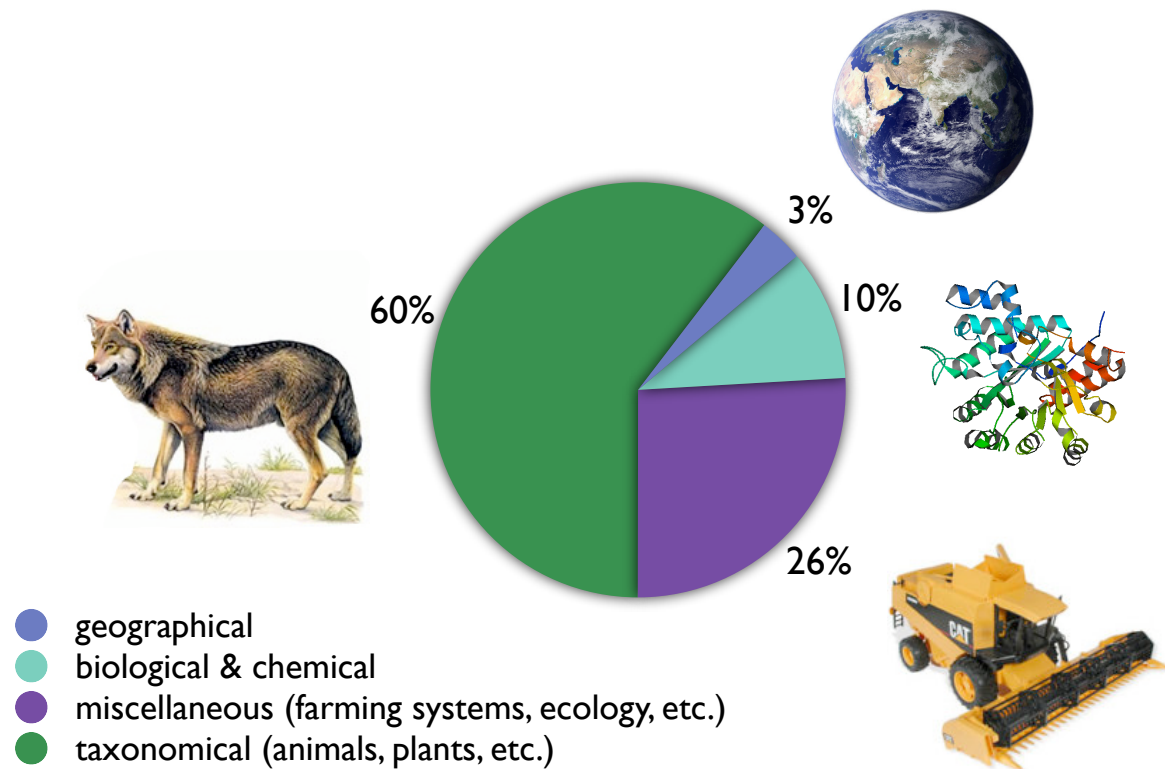
Precision



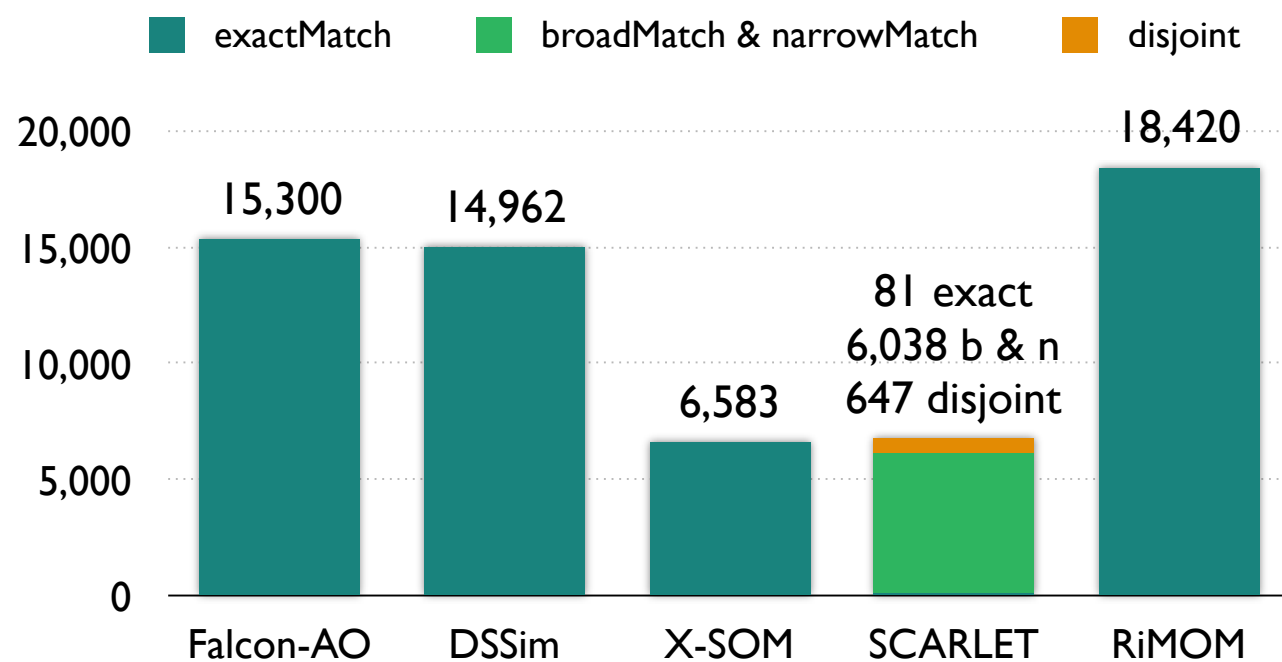
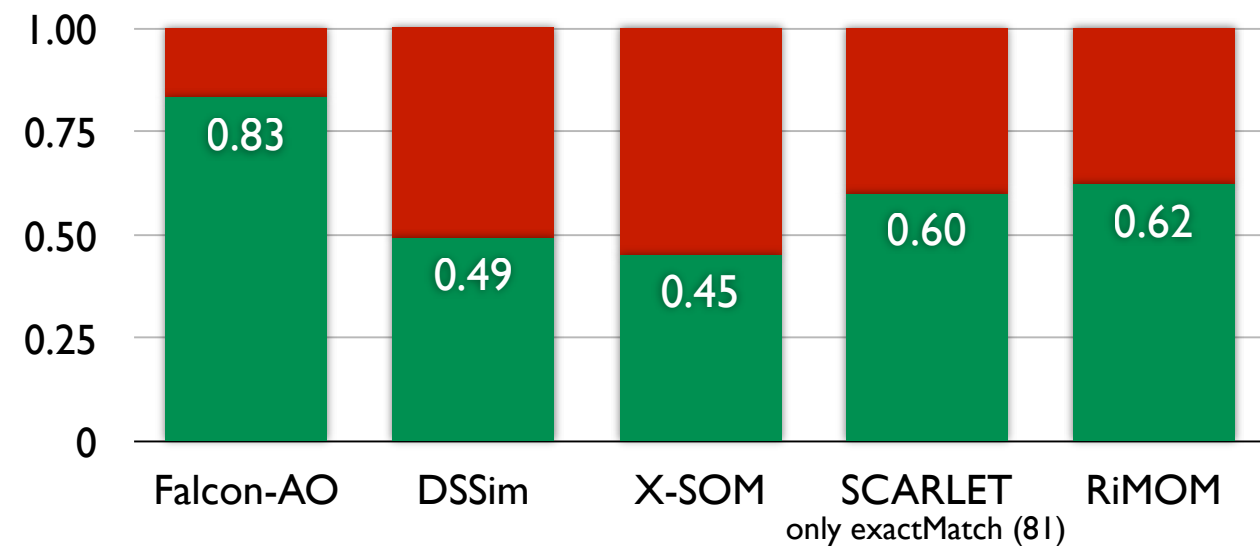
Recall



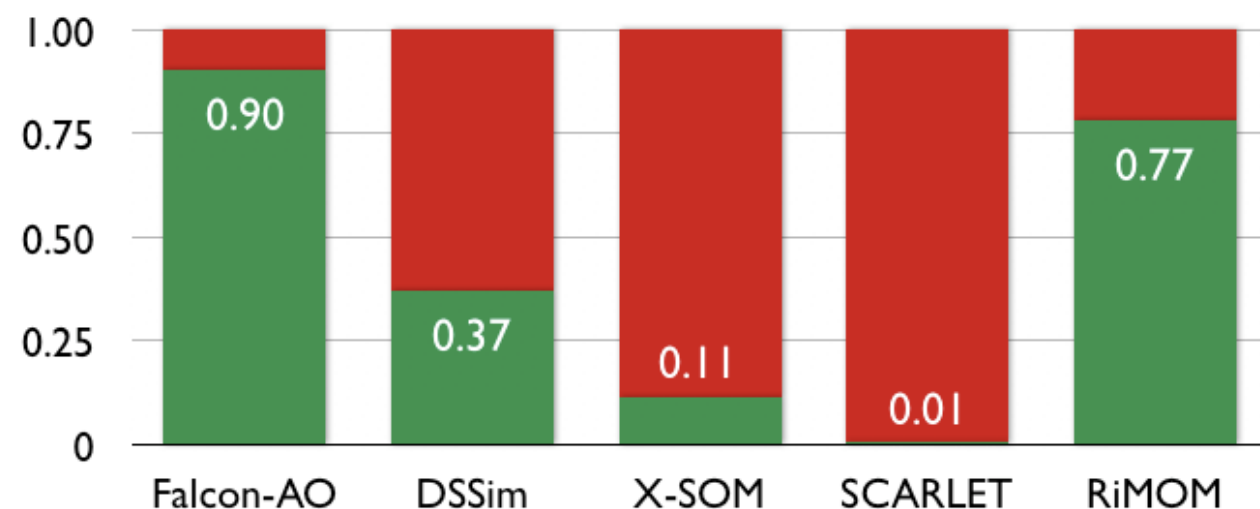
OAEI 2007 food task



Precision



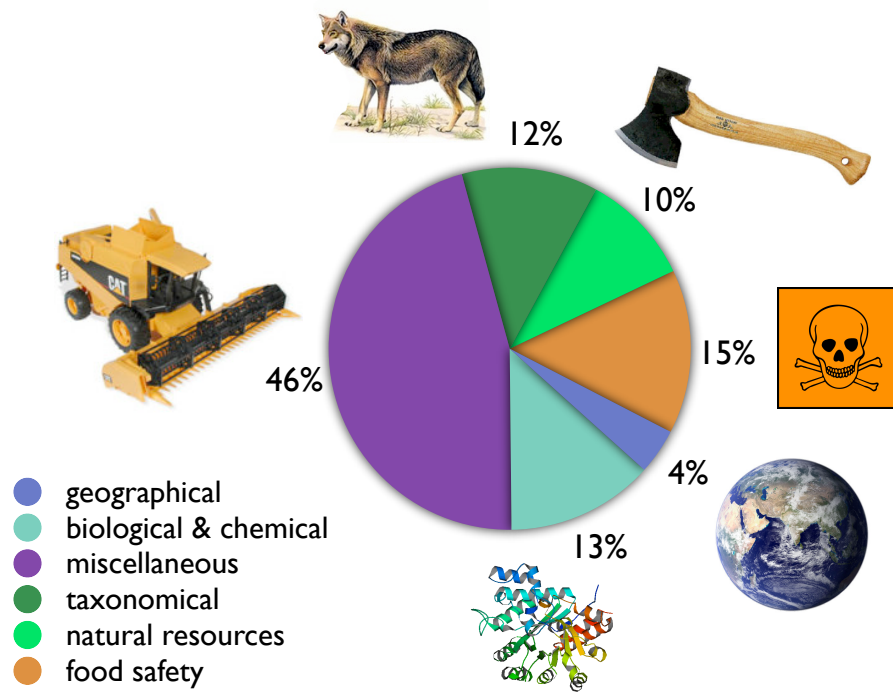
Recall



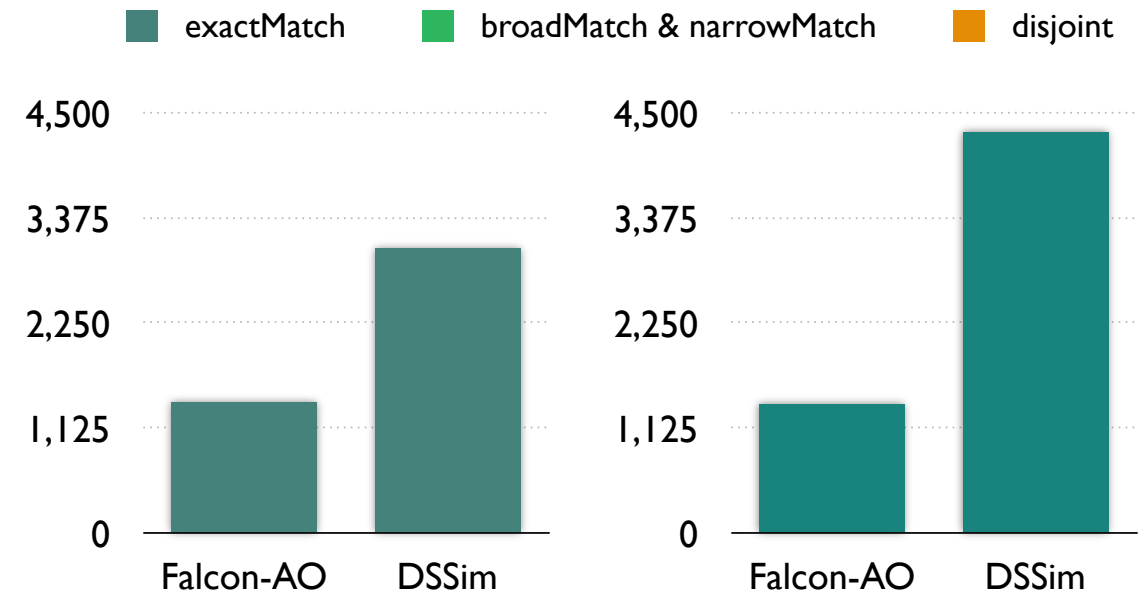
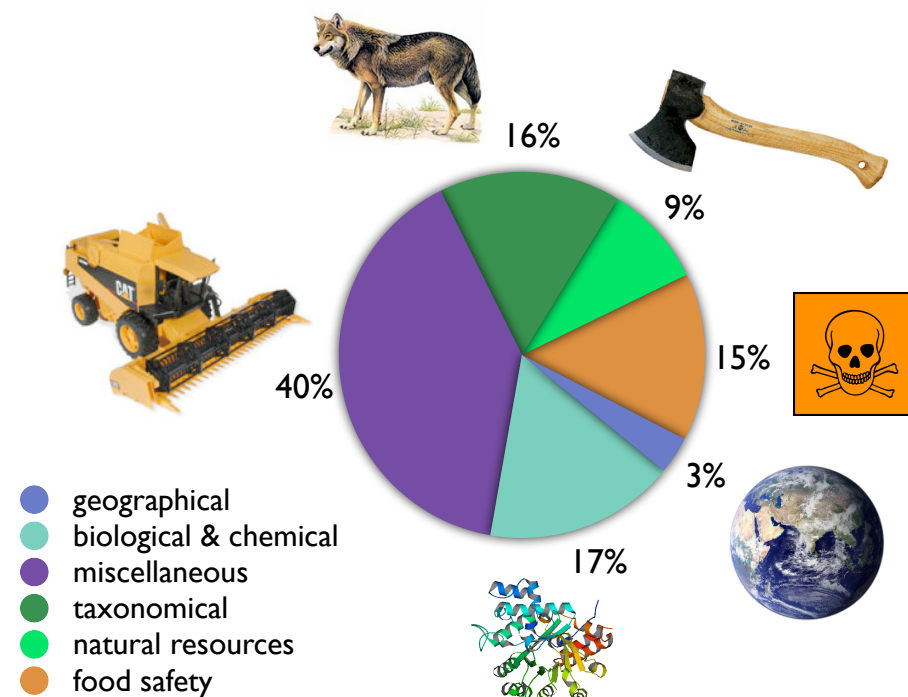
OAEI 2007 environment task



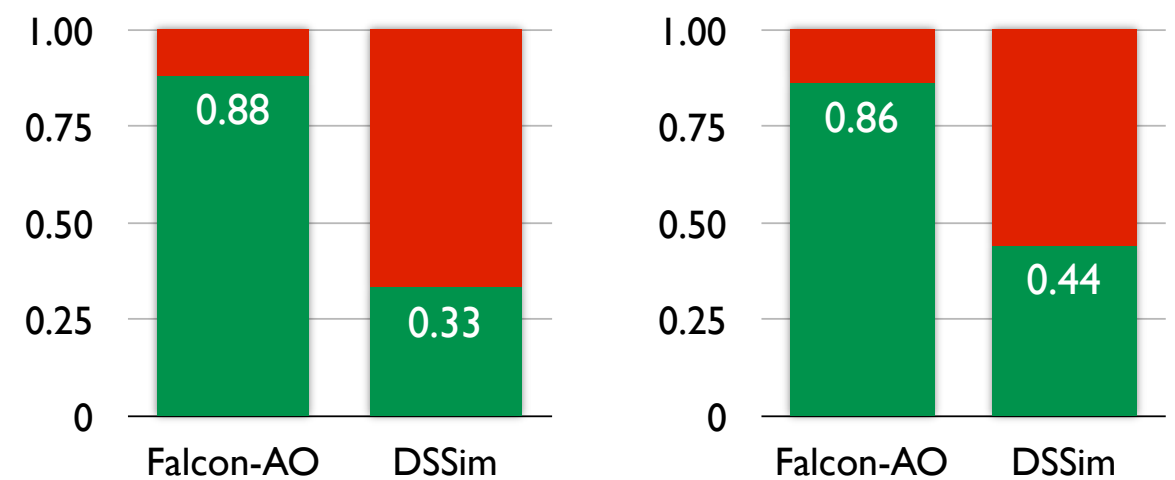
GEMET-AGROVOC



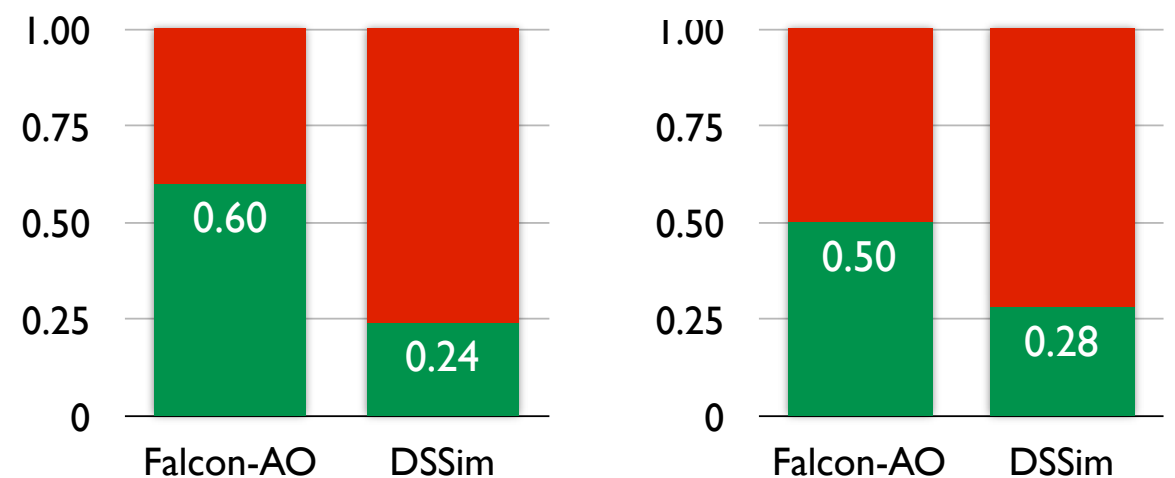
GEMET-NALT



Precision



Recall



GEMET-AGROVOC

GEMET-NALT

conclusions

- results improved significantly, especially in Recall, but interesting matches are still missing
- system design lessons learnt:
 - systems should **first** find the easy matches and then carefully extend to harder matches
 - systems should **only** try to find more matches when they do not already have a good match
 - systems should attempt to learn which lexical patterns hold in parts of the thesauri to distinguish “Bos taurus” < “Bos” from “lime stone” < “stone”
- systems should attempt to exploit background knowledge, alignment is really “AI-hard”

Evaluating Ontology Alignment Techniques

Willem Robert van Hage
VU University Amsterdam

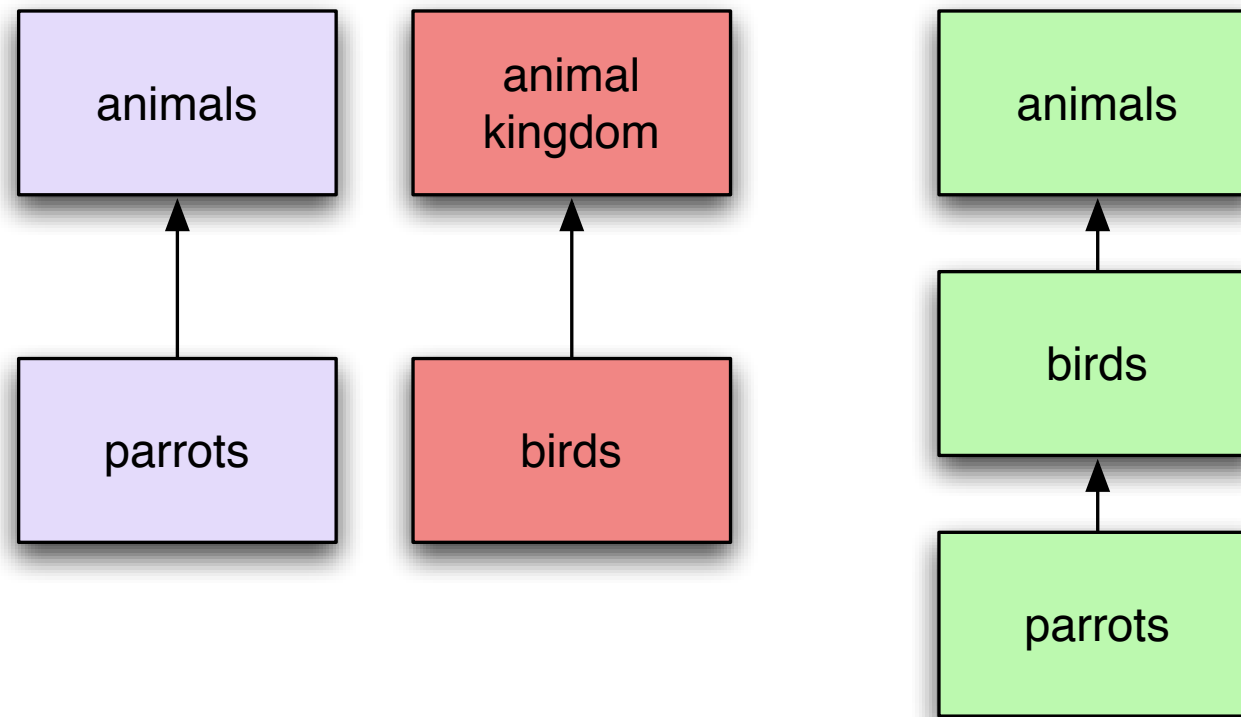
~~Evaluating~~ Ontology Alignment ~~Techniques~~ **why bother?**

Willem Robert van Hage
VU University Amsterdam

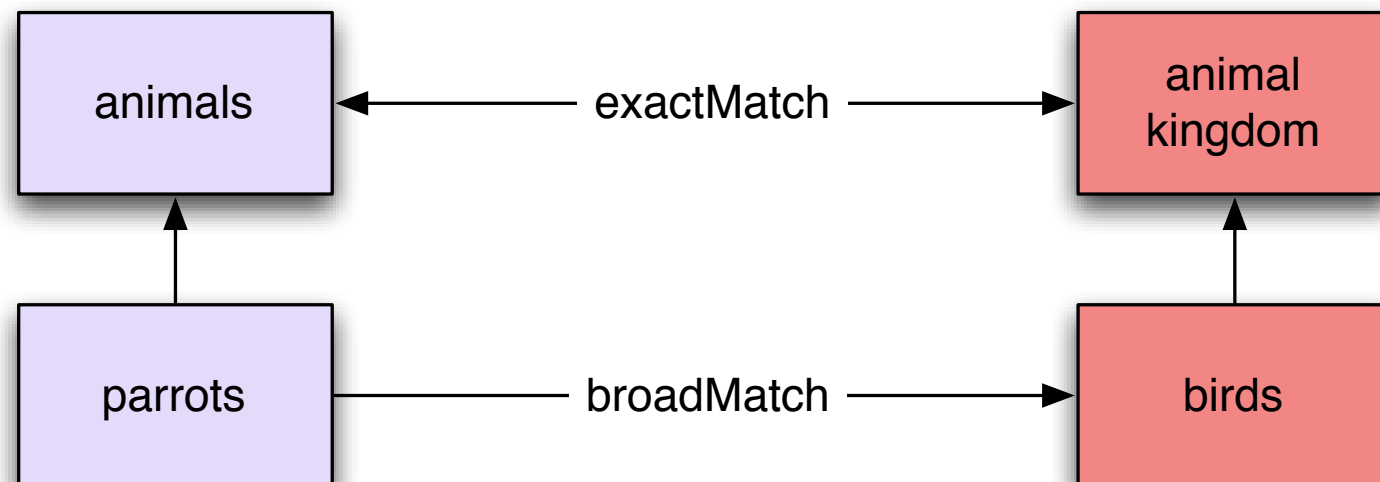
two approaches

- If you want to do information integration and you need to combine vocabularies you can do:
- **ontology merging**
 - start with two ontologies, end with one
 - merge some concepts, copy others, perhaps delete some
- **ontology alignment**
 - start with two ontologies, end with three
 - add relations between concepts, sometimes add intermediate concepts
 - two ontologies stay unchanged

merging



alignment



merging or alignment?

merging or alignment?

alignment.

why not merge?

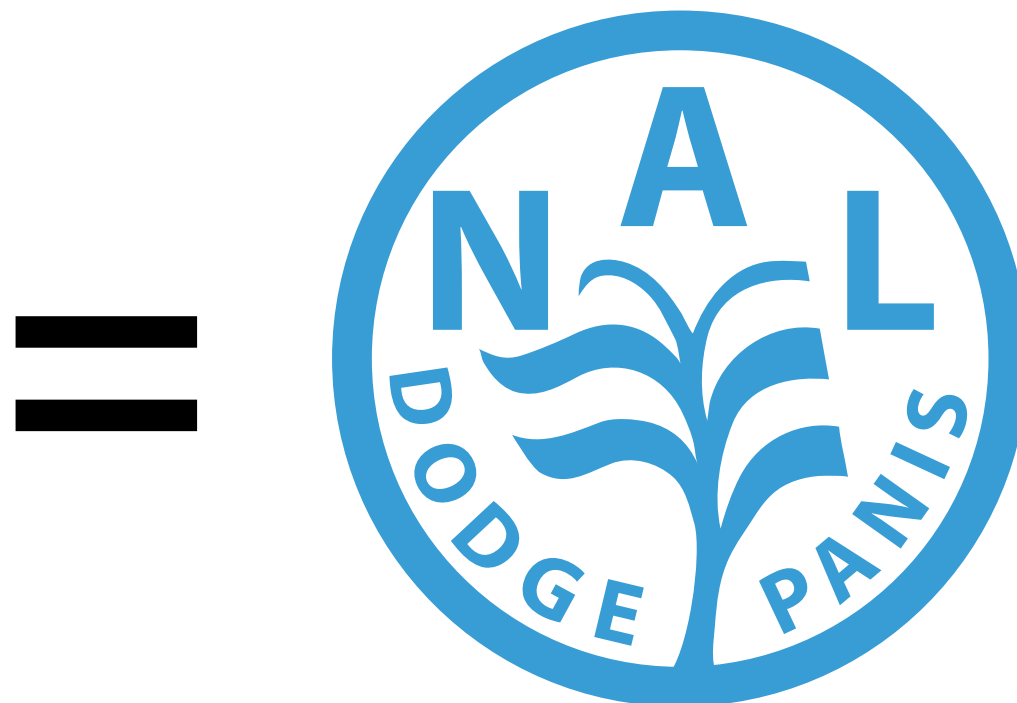
why not merge?



+



why not merge?



just kidding...

merging or alignment?

alignment.

merging or alignment?

alignment.

- alignment gives you more freedom to manage the combined resources in the future

merging or alignment?

alignment.

- alignment gives you more freedom to manage the combined resources in the future
- three important properties of ontology alignment:

merging or alignment?

alignment.

- alignment gives you more freedom to manage the combined resources in the future
- three important properties of ontology alignment:
 - I. the alignment itself is a separate collection

merging or alignment?

alignment.

- alignment gives you more freedom to manage the combined resources in the future
- three important properties of ontology alignment:
 1. the alignment itself is a separate collection
 2. alignment relations allow for subtle differences to be pointed out, but not removed

merging or alignment?

alignment.

- alignment gives you more freedom to manage the combined resources in the future
- three important properties of ontology alignment:
 1. the alignment itself is a separate collection
 2. alignment relations allow for subtle differences to be pointed out, but not removed
 3. the original thesauri can keep their own separate lives while applications can make combined use of them

issues with merging

issues with merging

- **legal issues**
 - who owns the result?
 - what about ownership of past and future versions?

issues with merging

- **legal issues**
 - who owns the result?
 - what about ownership of past and future versions?
- **maintenance issues**
 - who is allowed to change the resulting thesaurus?
 - who will pay for future modifications?

issues with merging

- **legal issues**
 - who owns the result?
 - what about ownership of past and future versions?
- **maintenance issues**
 - who is allowed to change the resulting thesaurus?
 - who will pay for future modifications?
- **security issues**
 - future changes might reveal confidential plans to parties

issues with merging

- **legal issues**
 - who owns the result?
 - what about ownership of past and future versions?
- **maintenance issues**
 - who is allowed to change the resulting thesaurus?
 - who will pay for future modifications?
- **security issues**
 - future changes might reveal confidential plans to parties
- **legacy issues**
 - software and internal policies will have to be adapted to deal with the new “world view”

different points of view

- alignment allows different points of view to coexist
- is that good or bad? – *it's better than bad, it's good!*
- you can always ignore the other perspectives, while you can benefit from them whenever you like “you never lose”
- sometimes it is very interesting to see where the meaning of concepts clash
- it is definitely good in cases where merging is politically impossible or cooperation is hard to organize
 - on the web this is very common

dealing with differences

- within a thesaurus mixed points of view should be avoided, but when you cooperate they are unavoidable
- you have to deal with them one way or another
 - you can sit together, work out who's wrong and update the ontologies
 - you can ignore the problem and not link to each other
 - you can describe the differences and decide how to deal with them whenever it becomes relevant

rdfs:subPropertyOf, skos:closeMatch, skos:broadMatch, etc.
(as opposed to owl:sameAs or owl:equivalentClass)

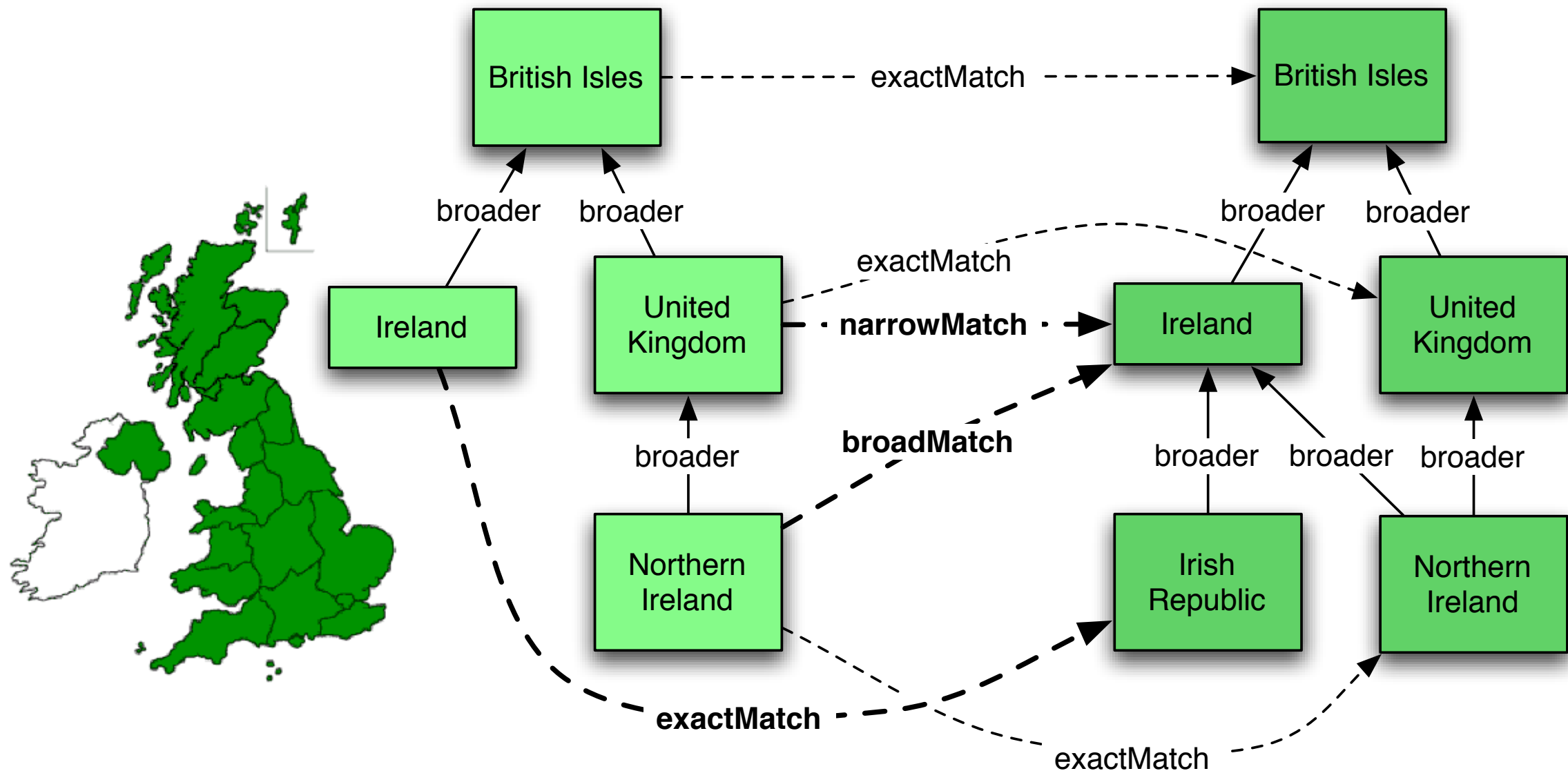
are your friends

example

let's think about the consequences

AGROVOC

NALT

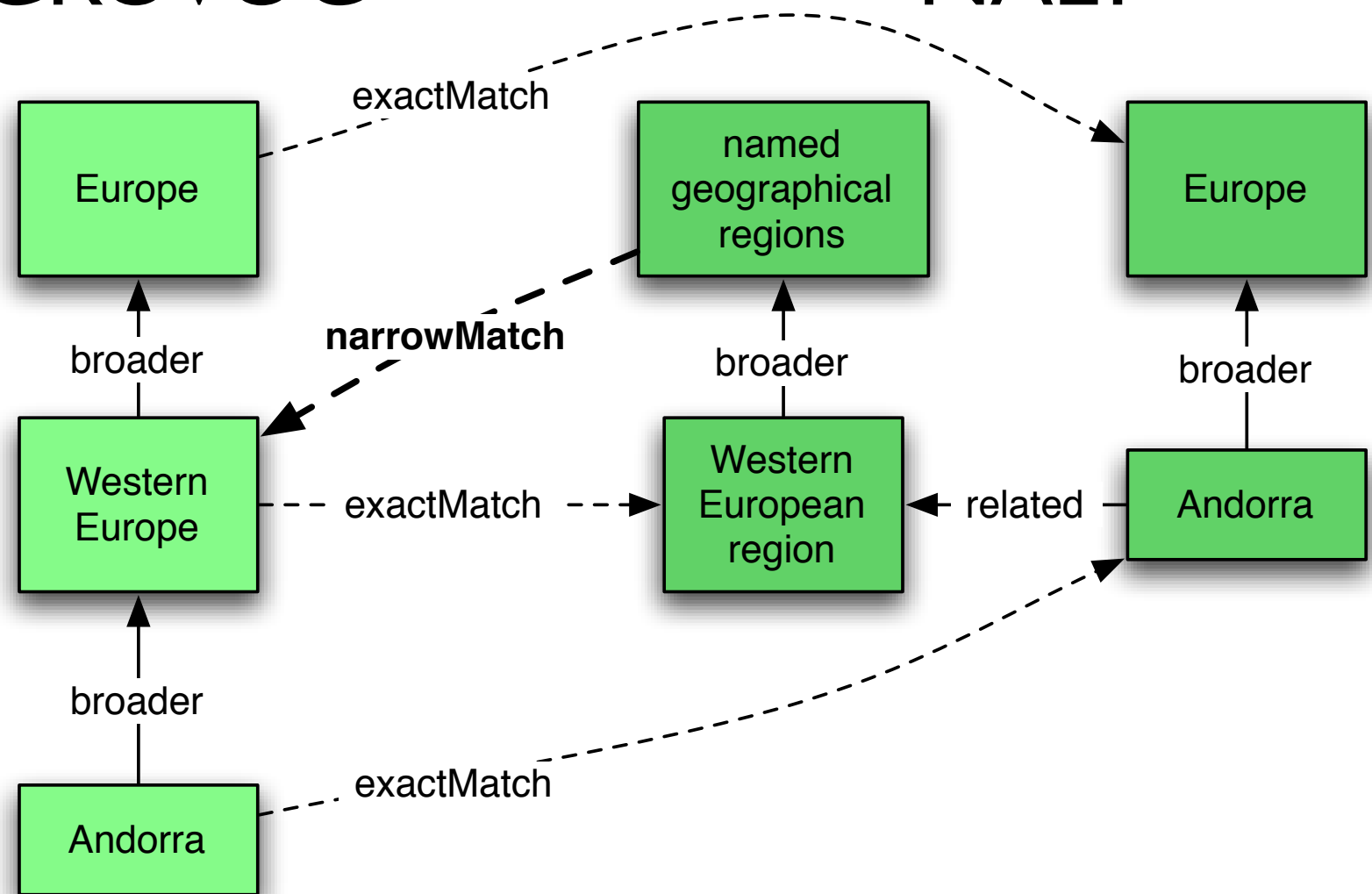
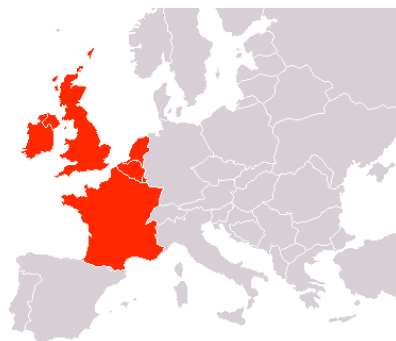
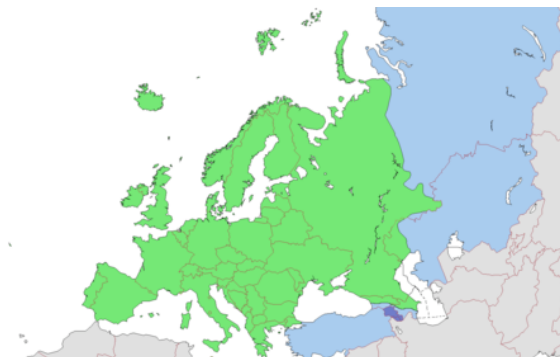


example

let's think about the consequences

AGROVOC

NALT



a final remark

about power and the web

- in the past you gained the most power by constraining access to your information
- now you can also gain power by having people use your information and extend it for you
 - sharing makes you a de facto authority: people use whatever works and is available
 - sharing makes others do part of your work for you: when other people openly link their information to yours you can also make use of the link
- consider benefitting from publishing linked data by making it or by aligning with it

