OAEI 2006 - food task

Willem Robert van Hage

Free University Amsterdam
and TNO Science & Industry

NALT & AGROVOC

NAL
Thesaurus

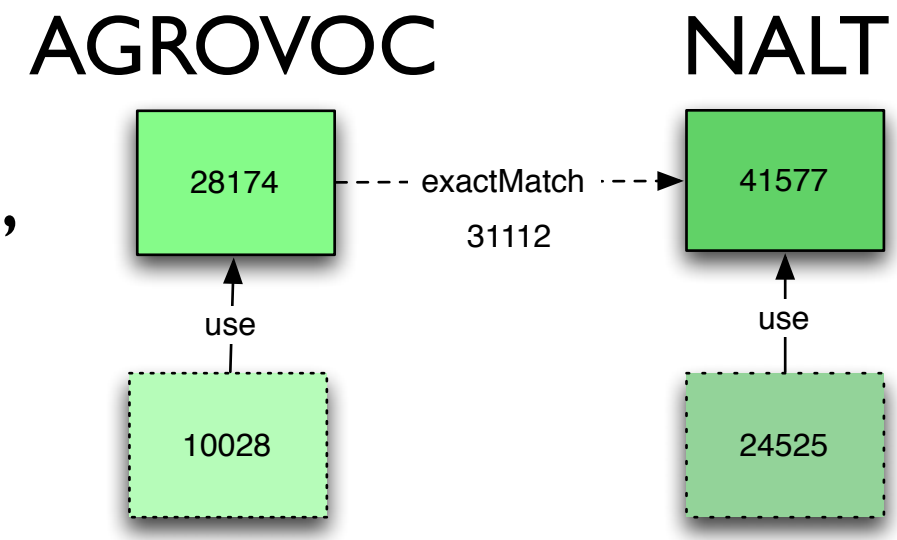
AGROVOC
Thesaurus

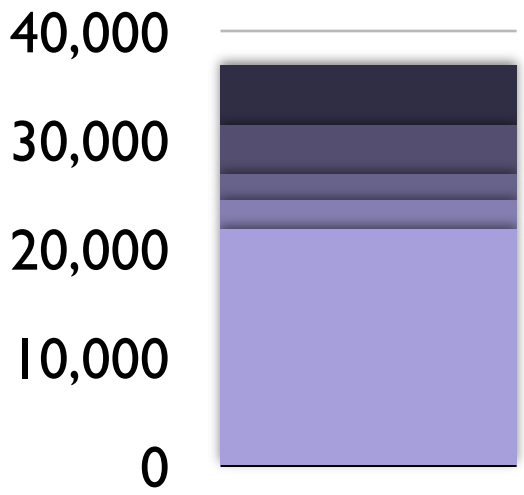
AGRICOLA
AgNIC
FSRIO

AGRIS
CARIS

some numbers

- 31112 mappings, all exactMatch



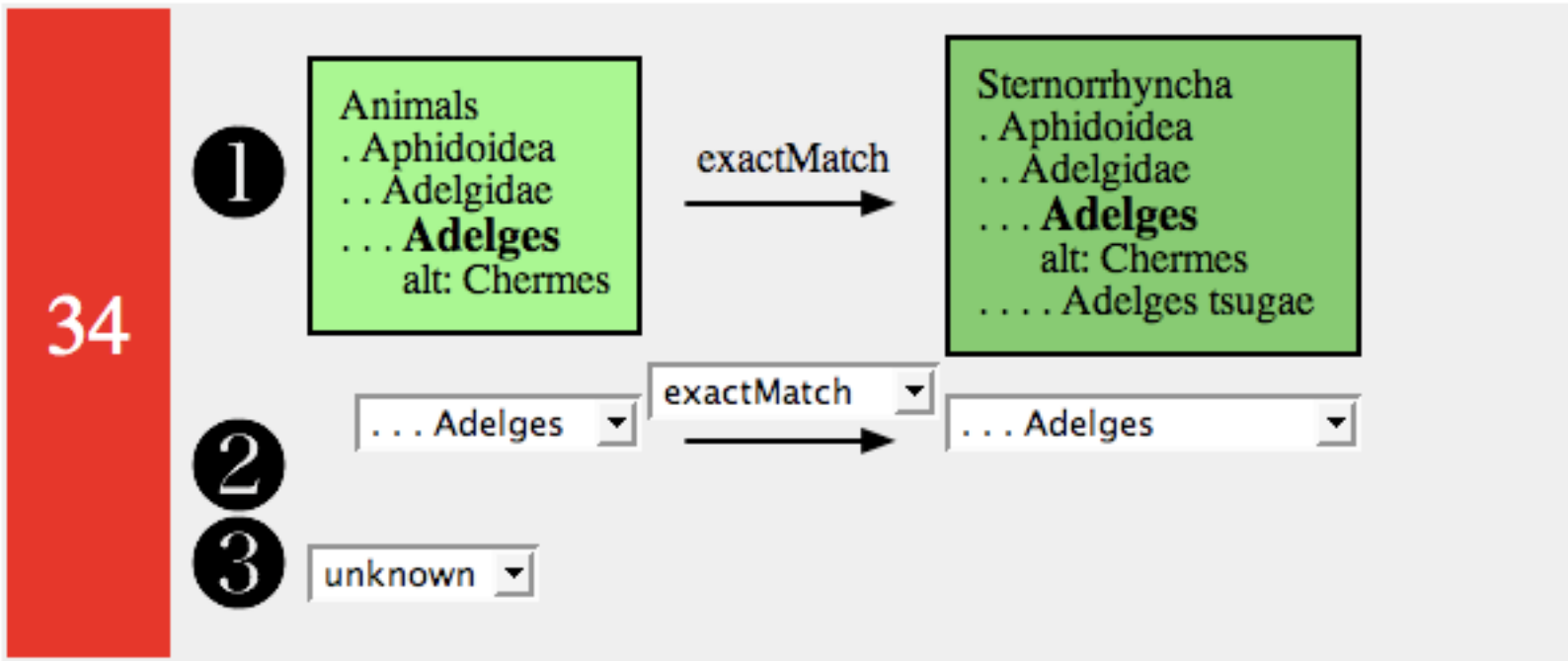


submitted by 1 participant
submitted by 2 participants
submitted by 3 participants
submitted by 4 participants
submitted by all participants

sample evaluation of precision

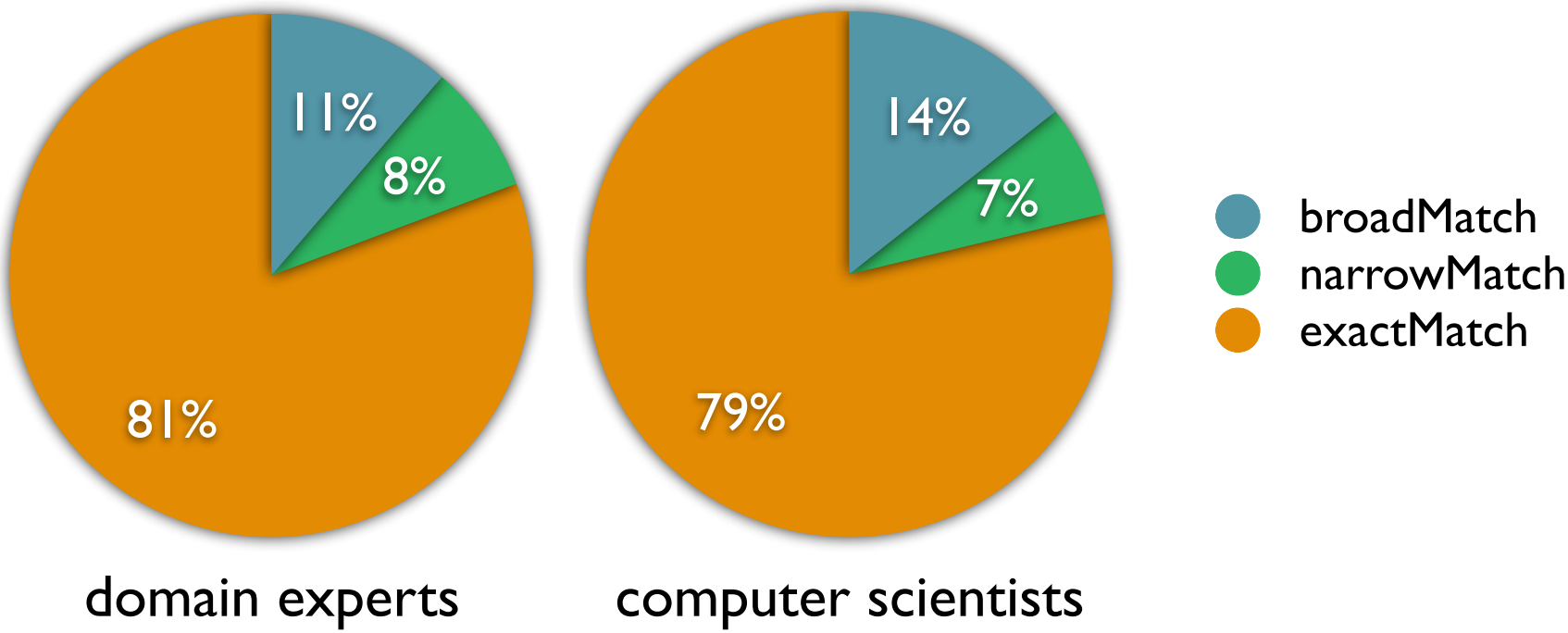
- domain experts evaluated 7% of the mappings
- these were chosen in such a way that
  - all participants were equally represented
  - all levels of consensus amongst the participants were equally represented
  - different topics were represented fairly
- evaluation repeated with computer scientists

sample evaluation of precision



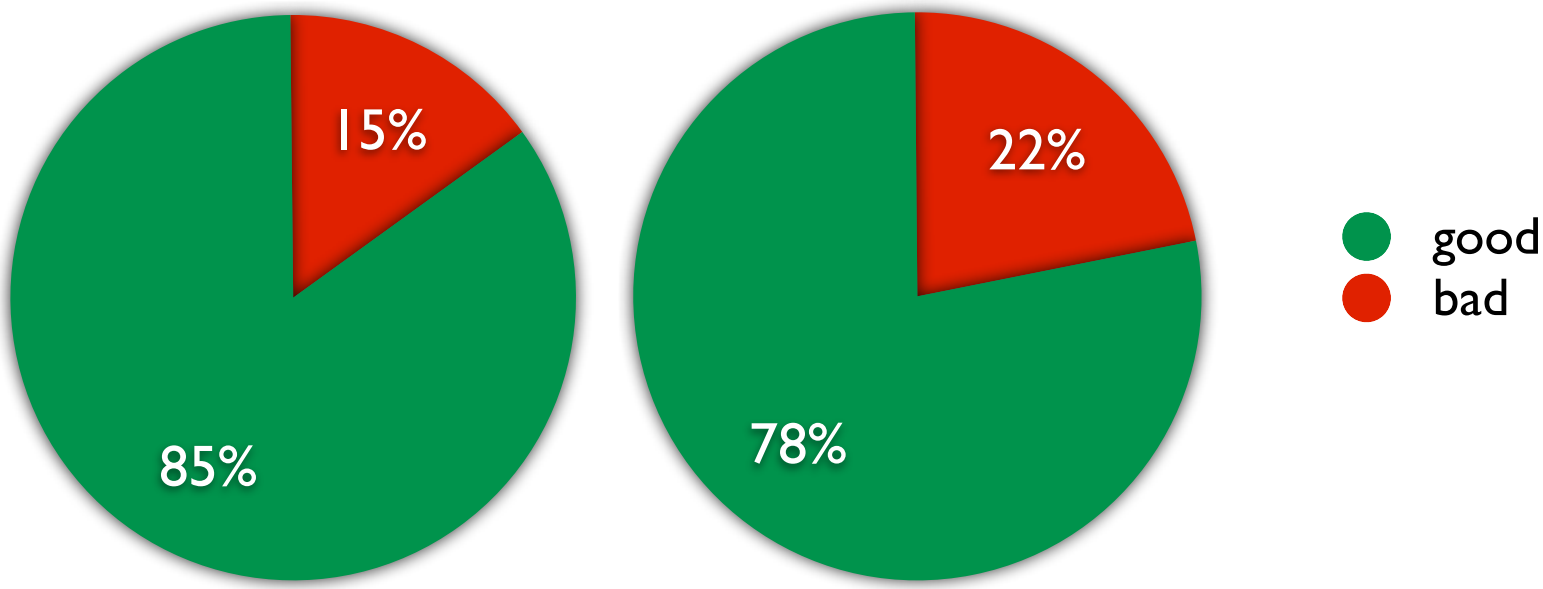
domain experts vs computer scientists

- 72% agreement between domain experts and computer scientists

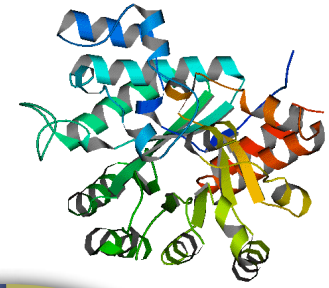


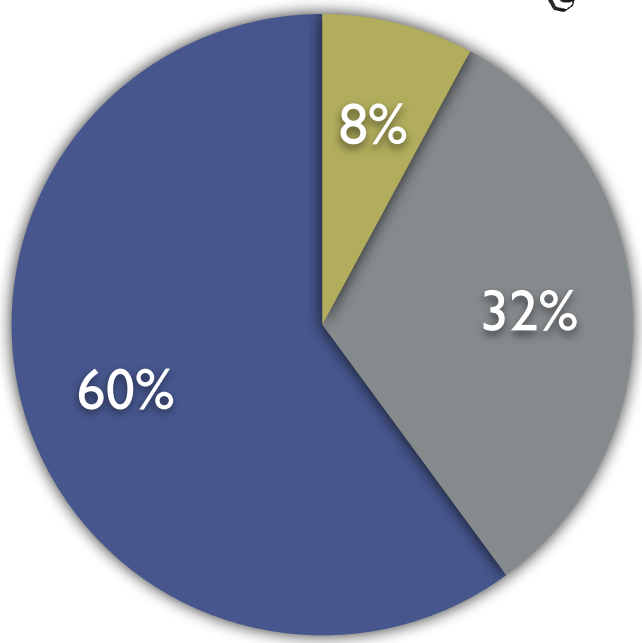
domain experts vs computer scientists

- 72% agreement between domain experts and computer scientists



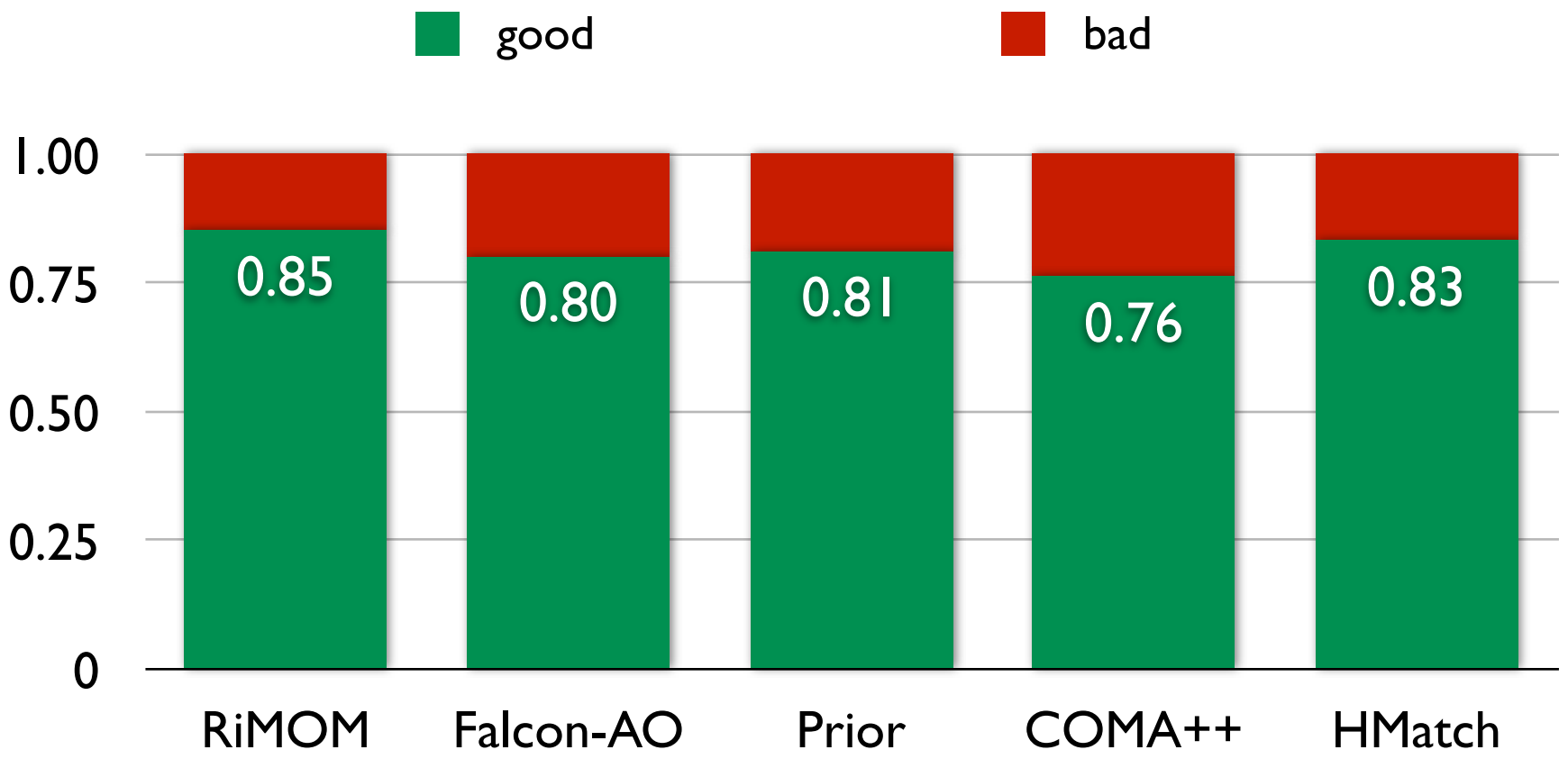
topics in the results





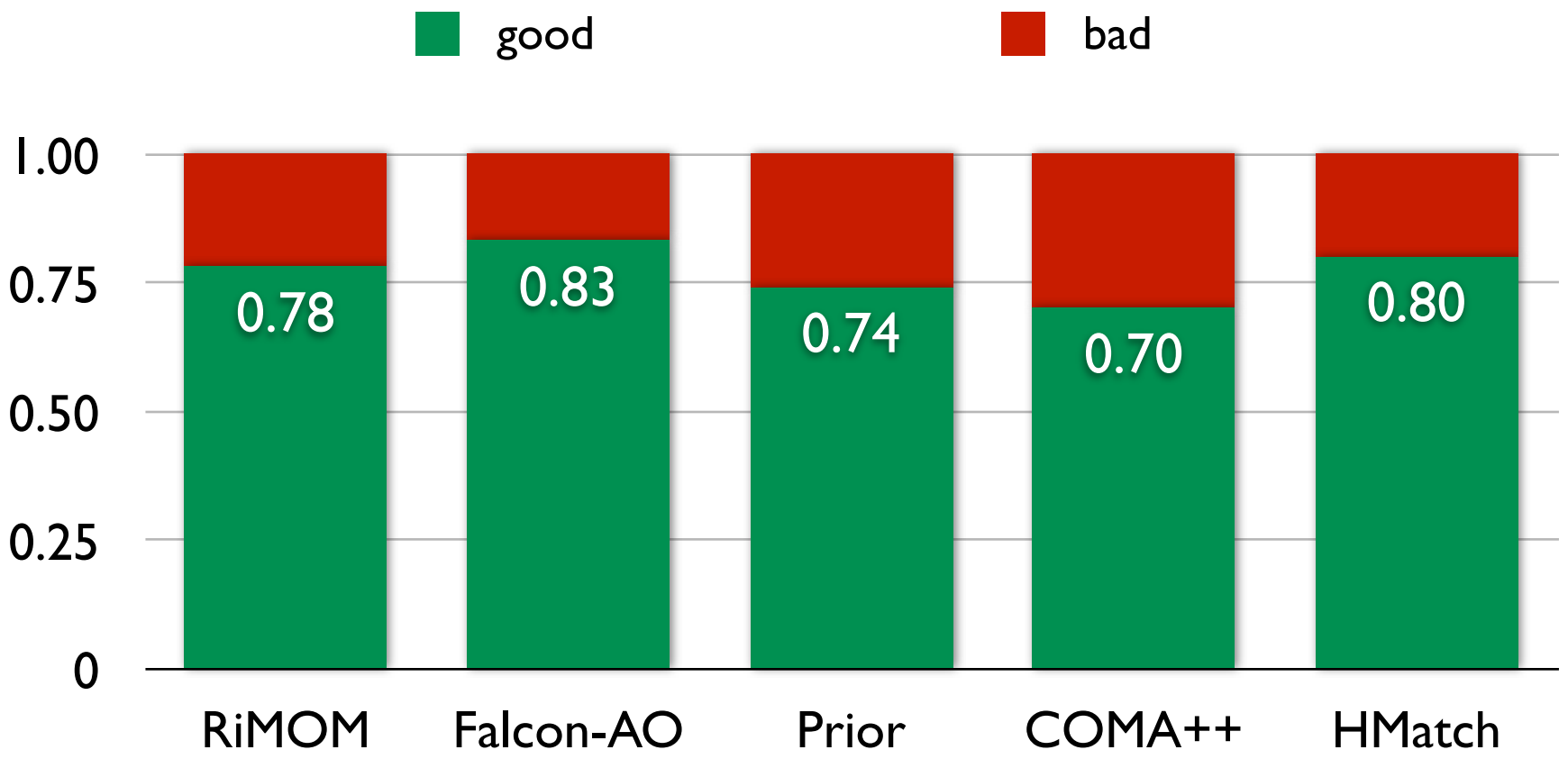
- biological & chemical
- miscellaneous (geographical, legislation, food stuffs, etc.)
- taxonomical

precision of biological & chemical mappings



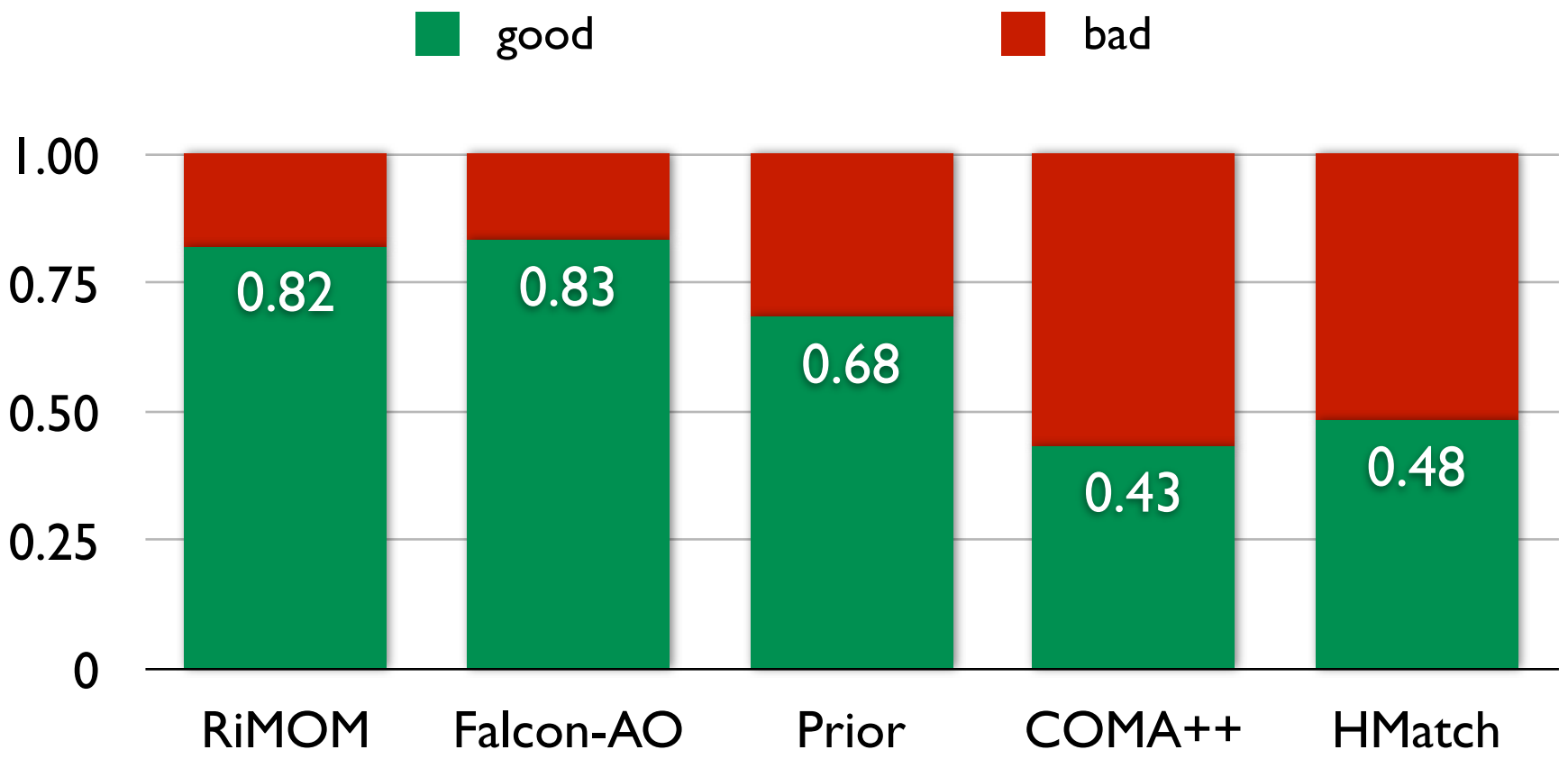
very specific terms, many spelling variants
different terms may mean the same thing

precision of miscellaneous mappings



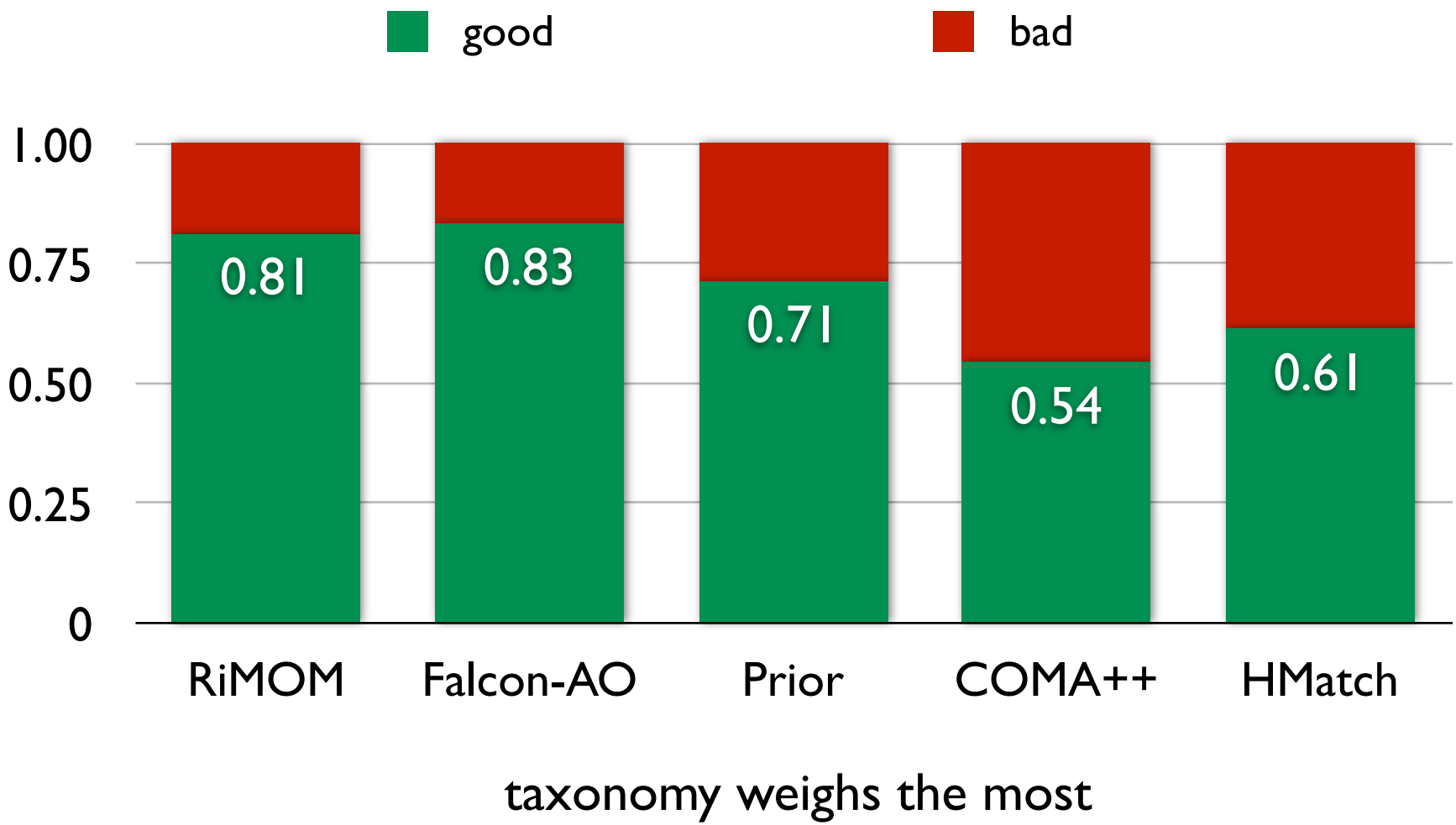
diverse topics, similar terms usually have similar meaning

precision of taxonomical mappings



many similar terms with different meaning
e.g. Canis Familiaris, Canis Lupus

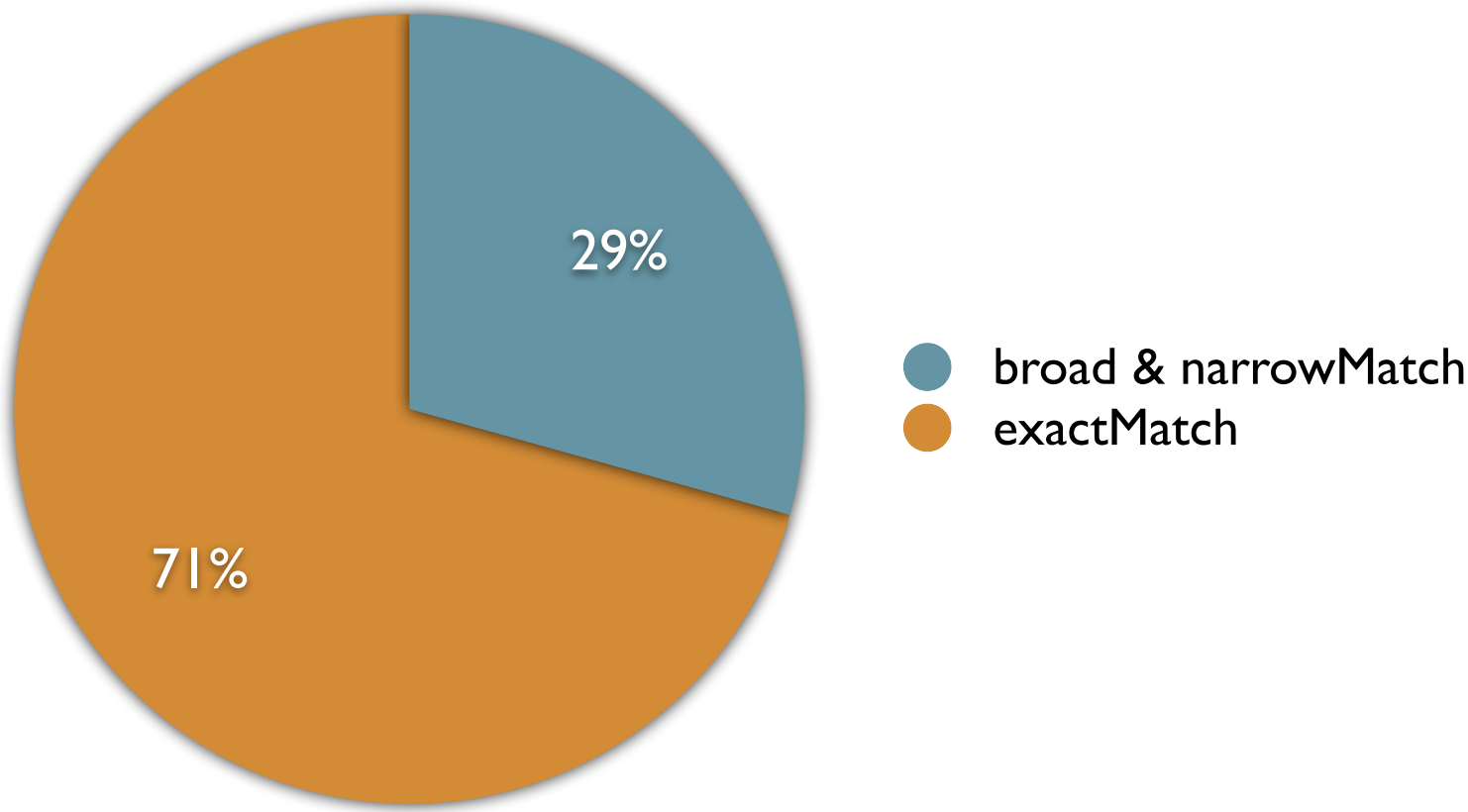
overall precision



sample evaluation of recall

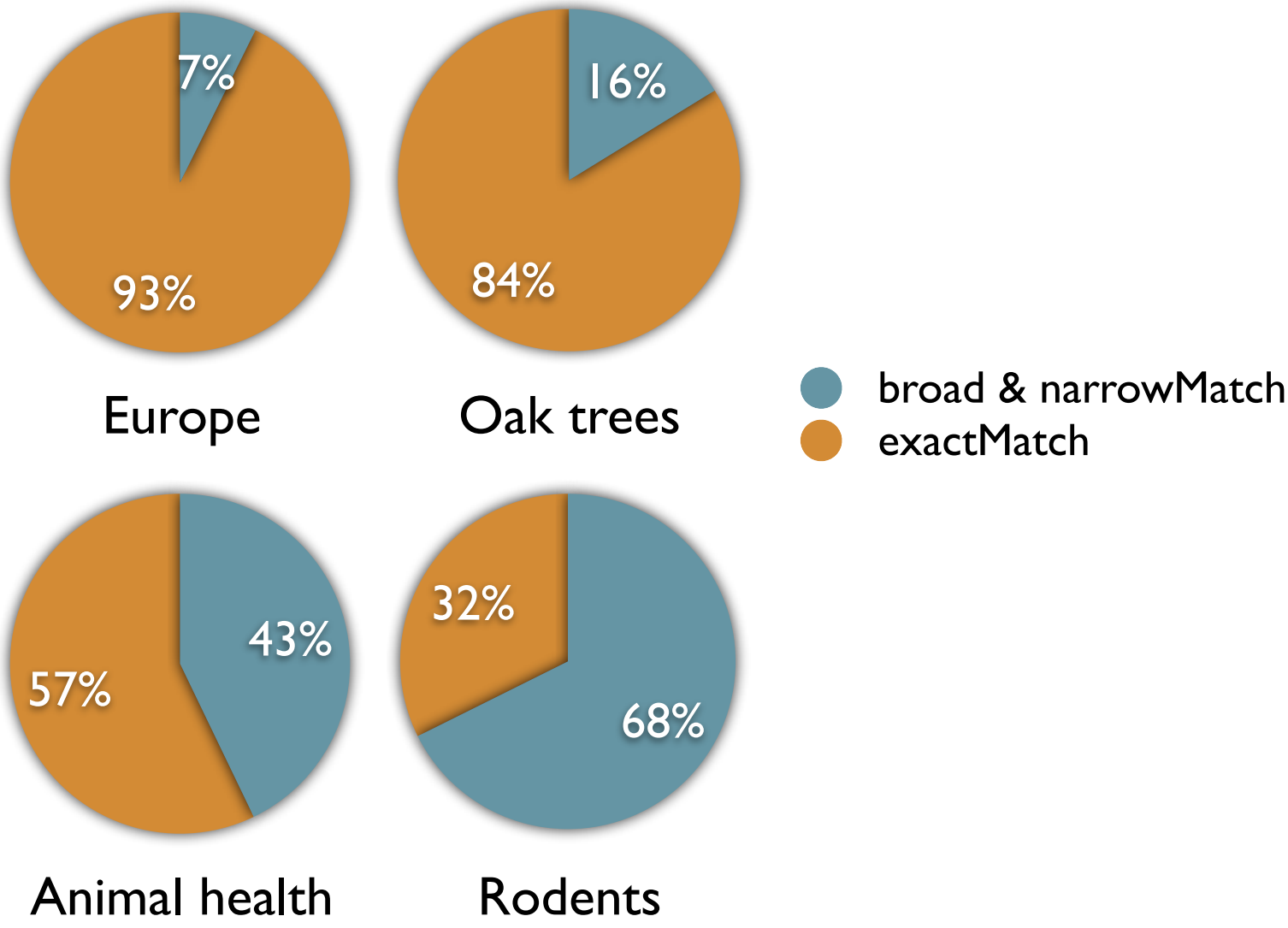
- evaluating a sample large enough to lead to significant results for recall was not feasible
- preliminary results are on 191 mappings in the miscellaneous and taxonomical domain
  - two sets from the miscellaneous domain: “animal health” and “geography of europe”
  - two sets from the taxonomical domain: “rodents” and “oak trees”

gold standard mapping relations

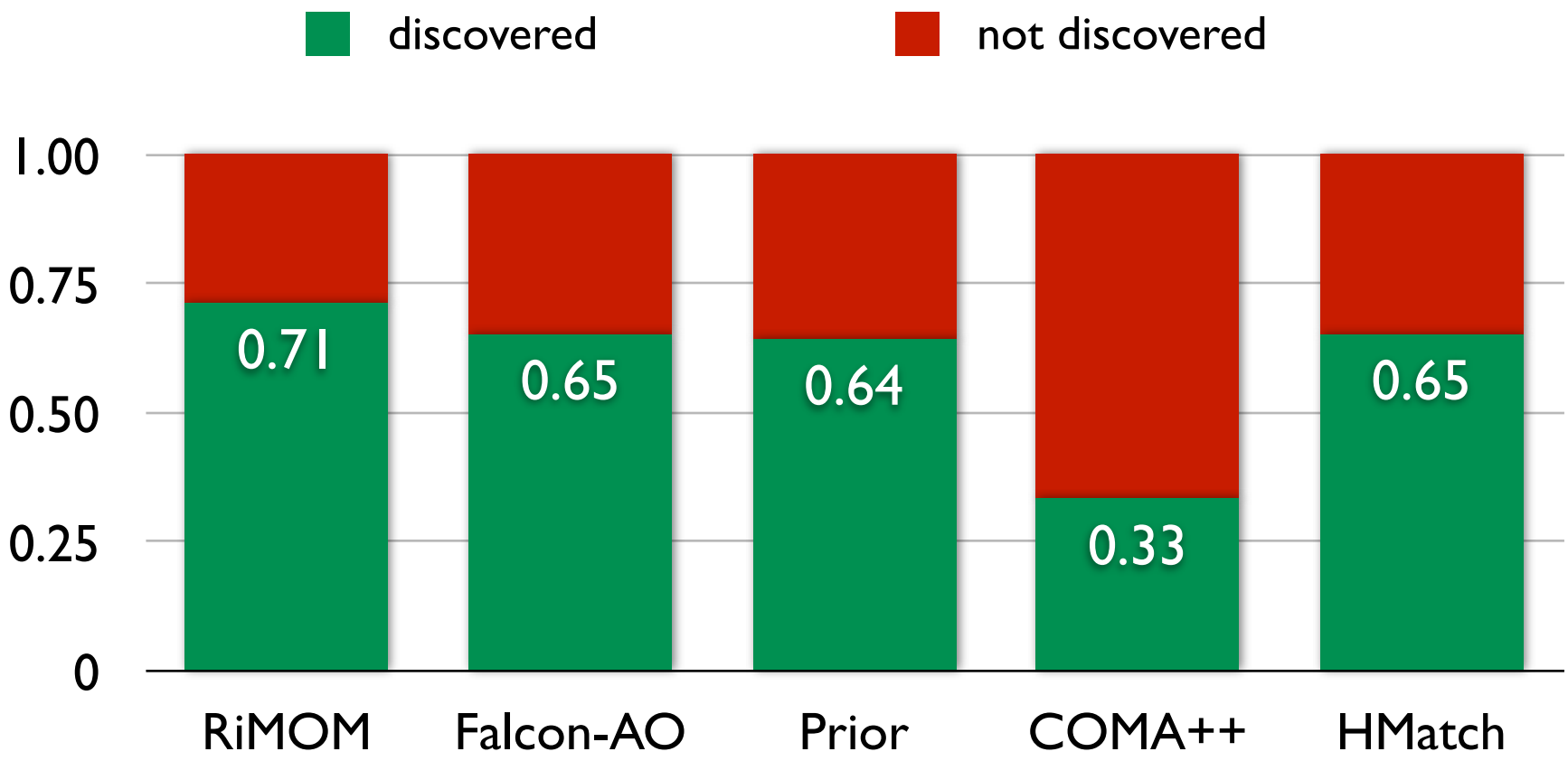


all topics

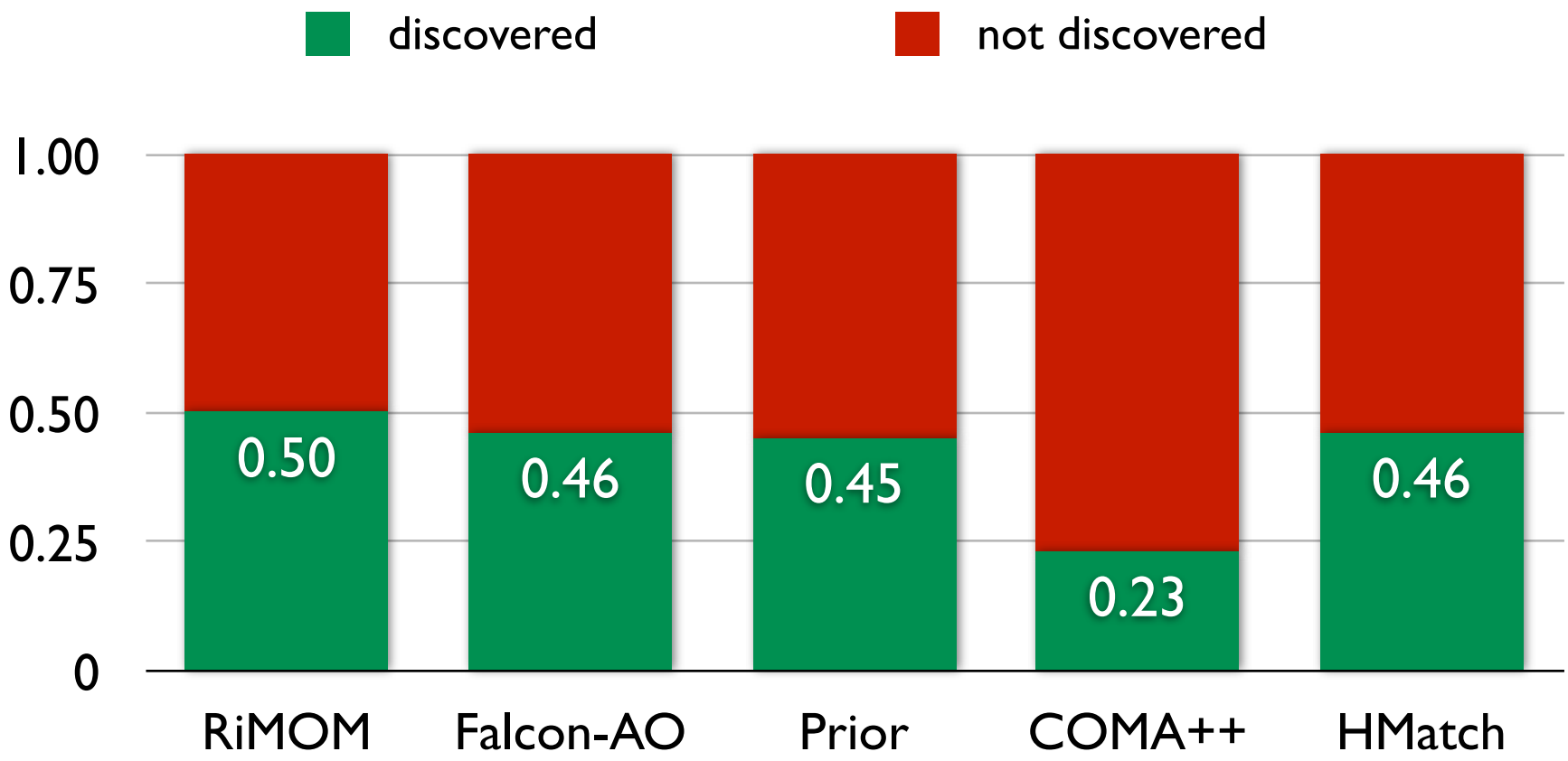
gold standard relations



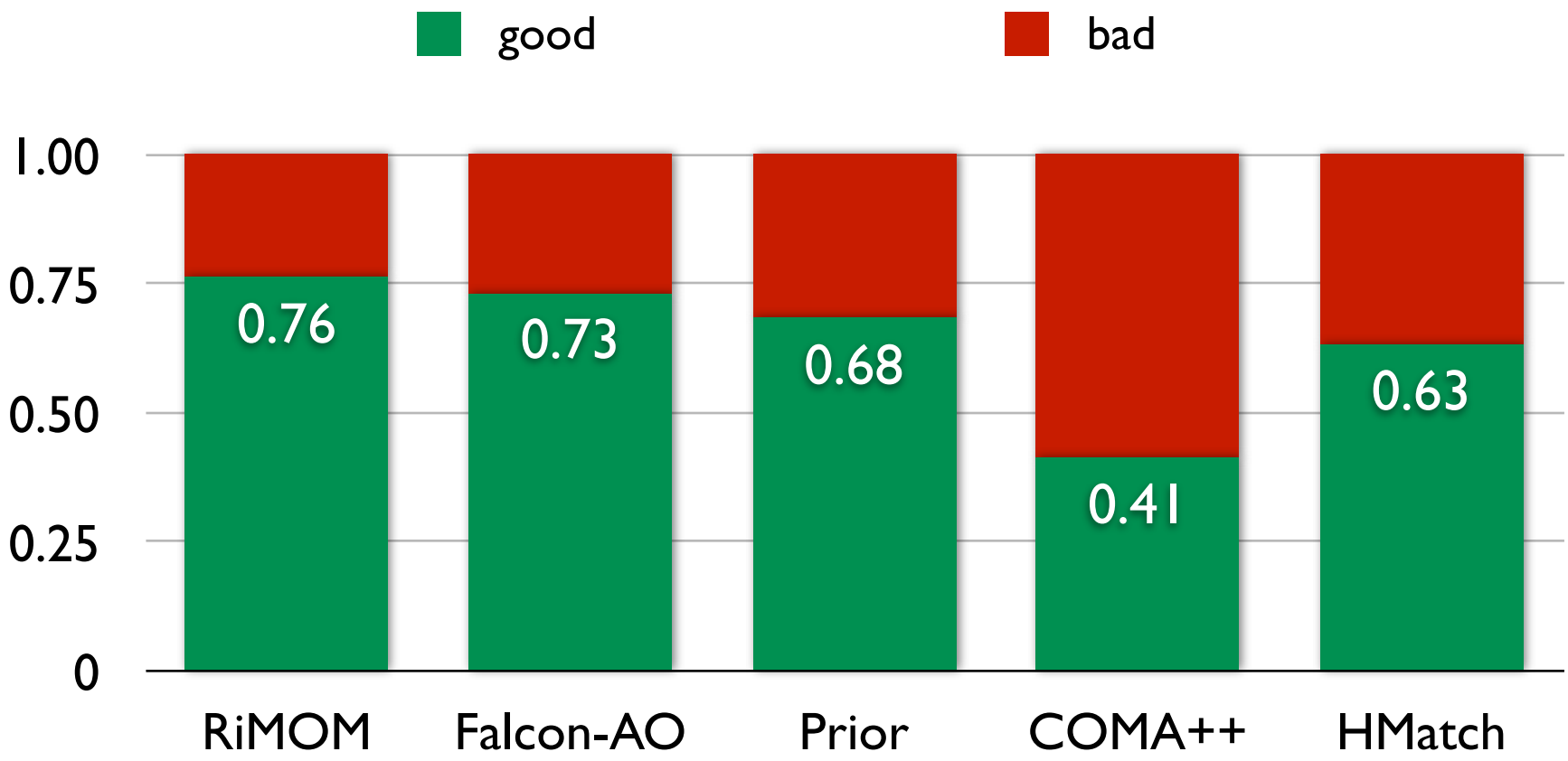
recall of only exactMatch mappings



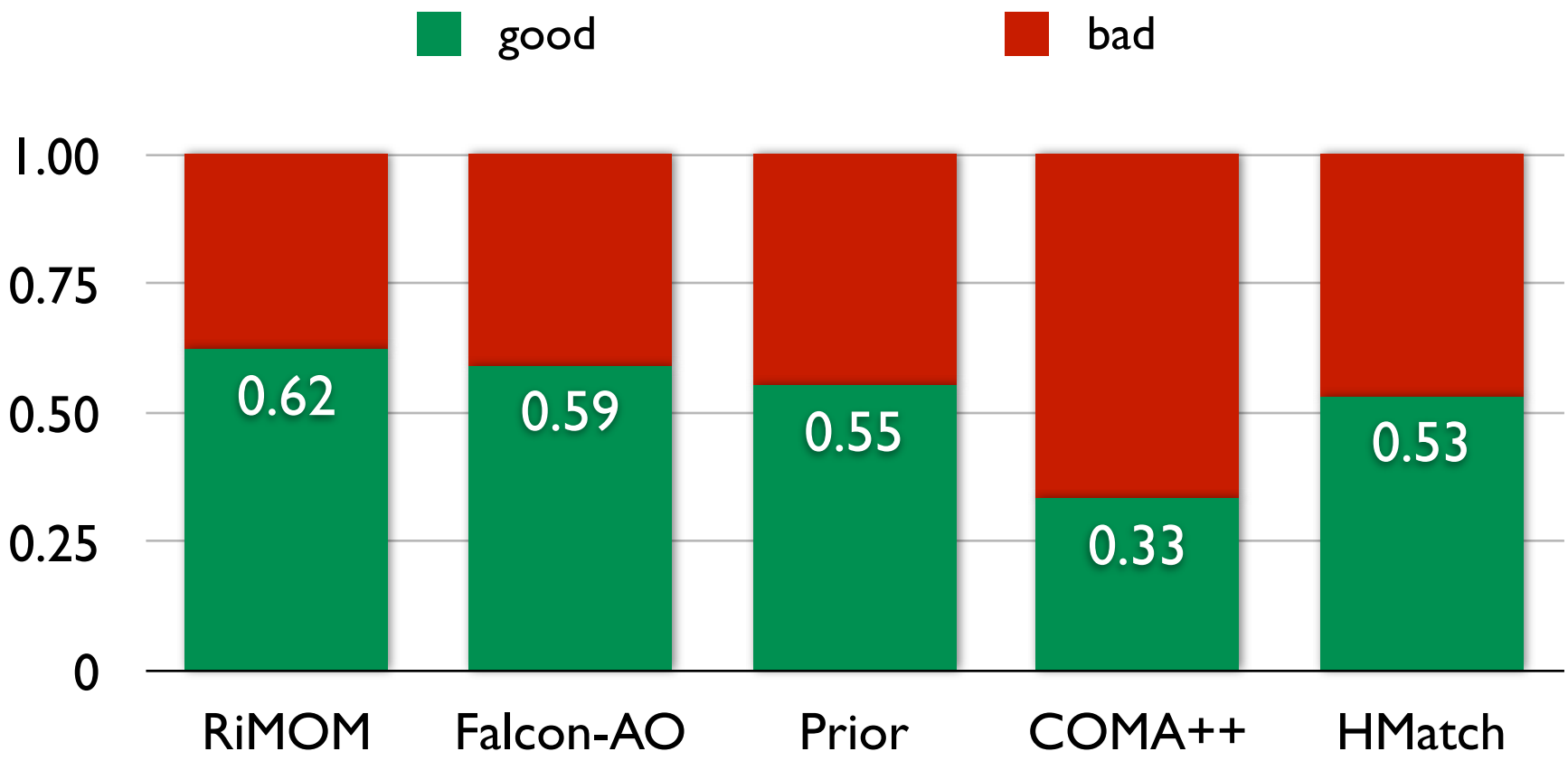
recall of
exact, broad & narrowMatch



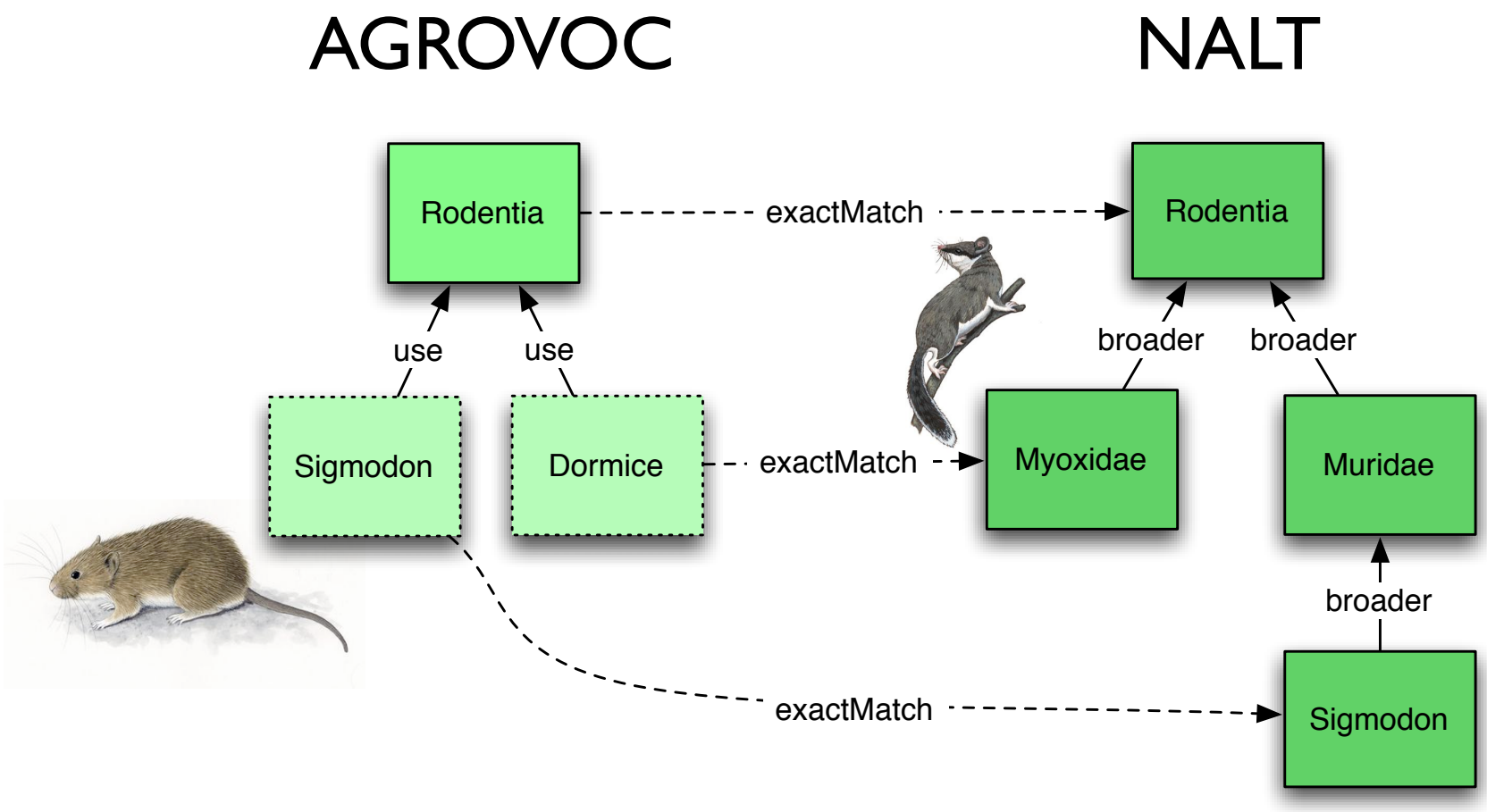
f-score of only exactMatch mappings



f-score of exact, broad & narrowMatch

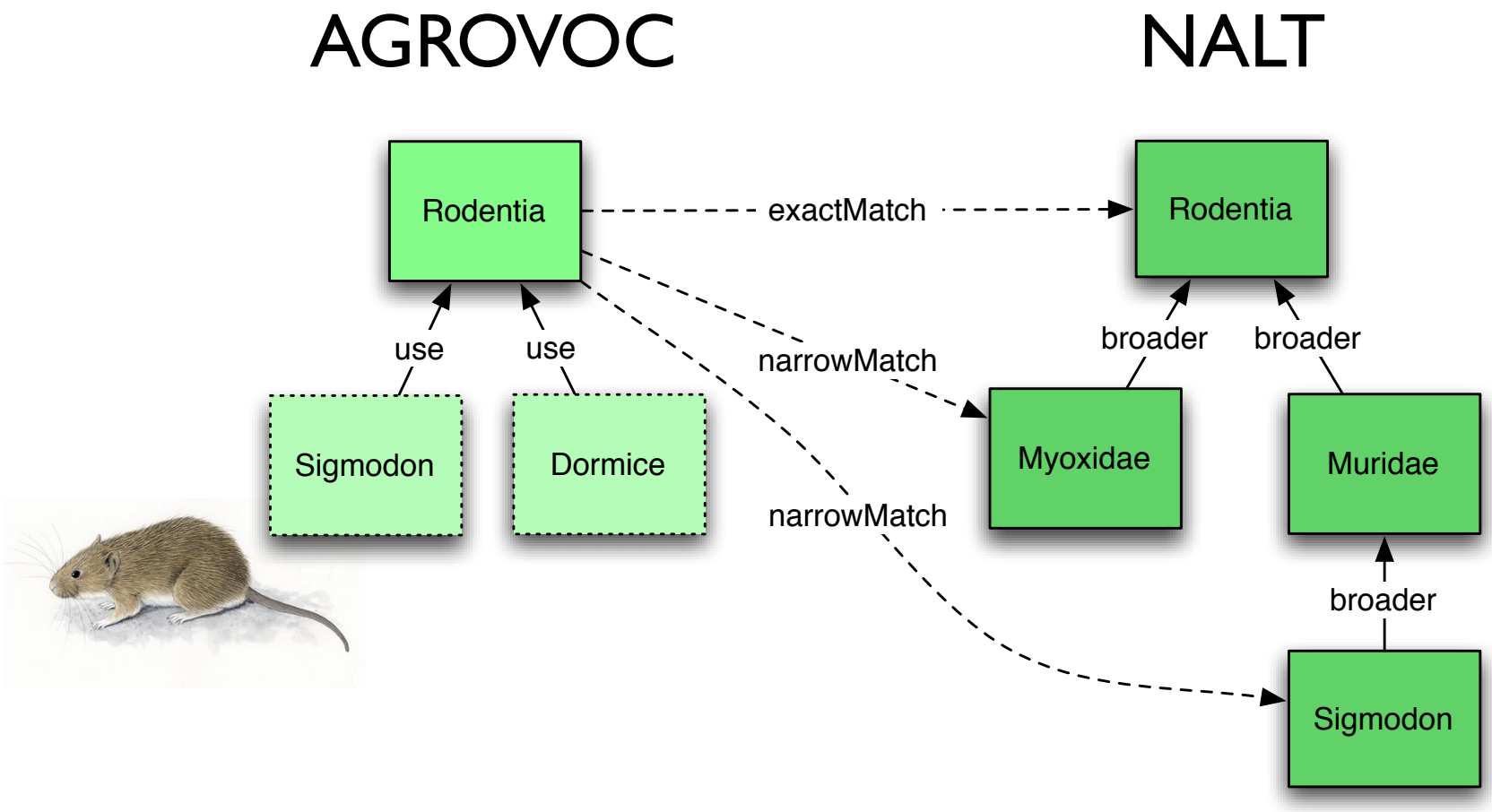


issues



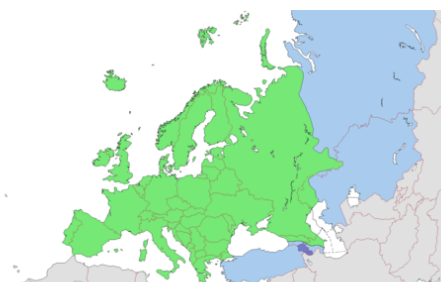
How do you deal with the USE relations / skos:altLabel?

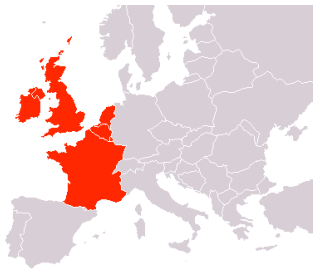
issues

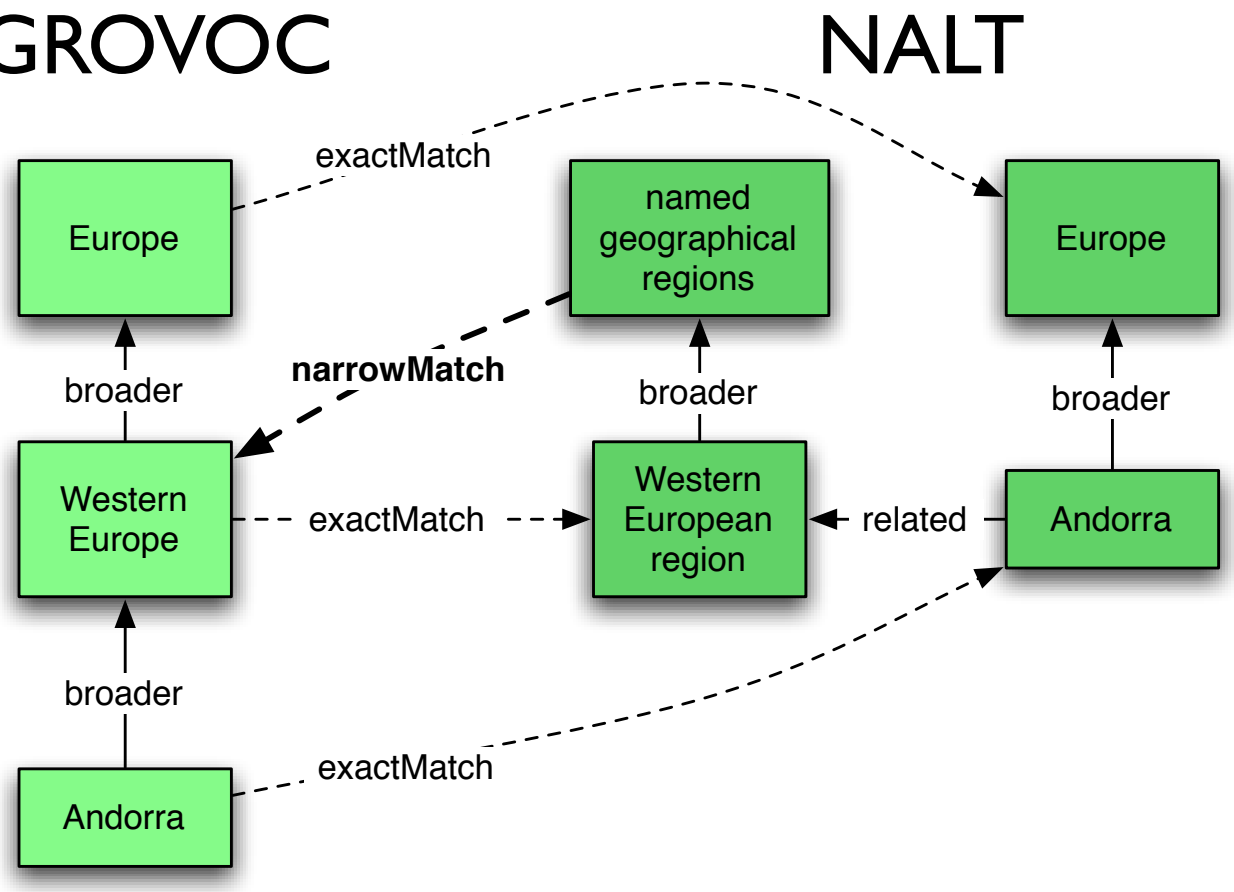


How do you deal with the USE relations / skos:altLabel?

issues

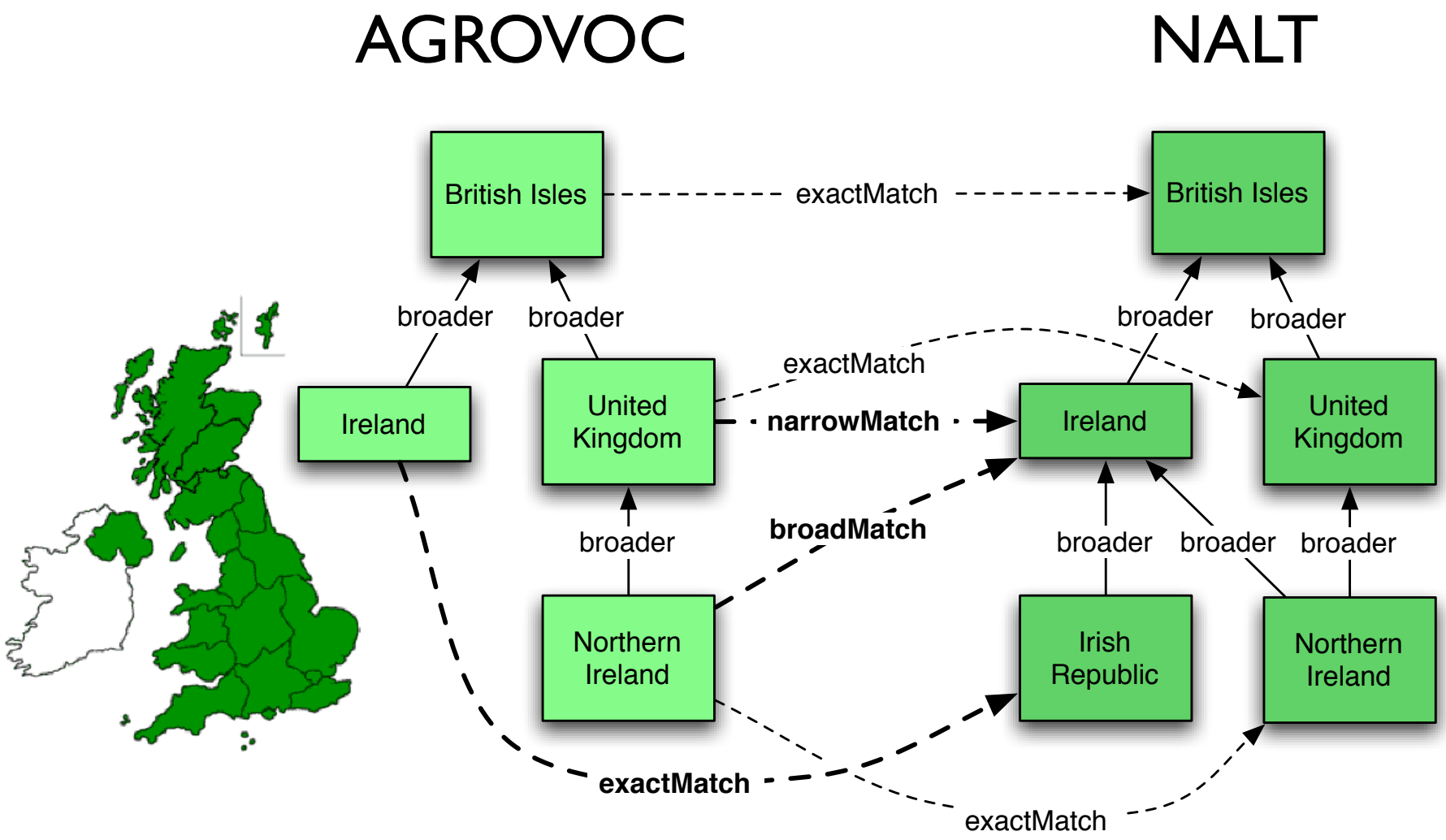






How do you find this relation?

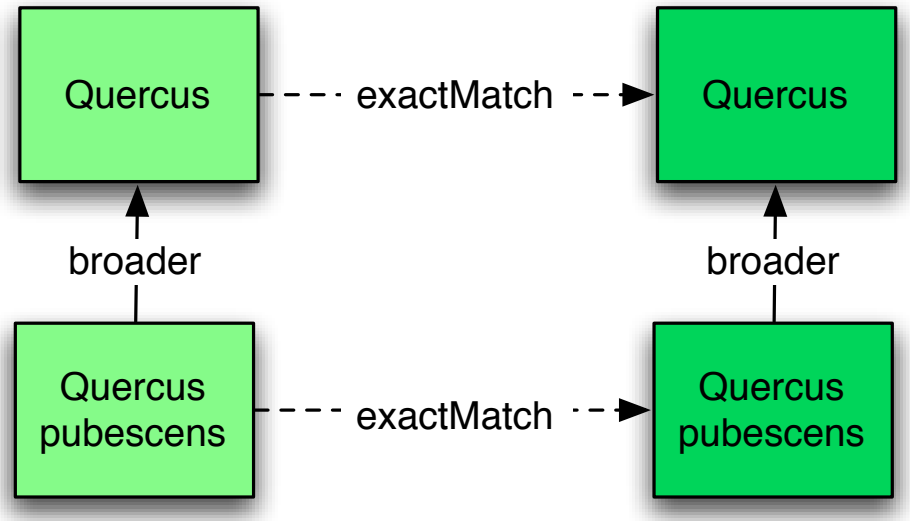
issues

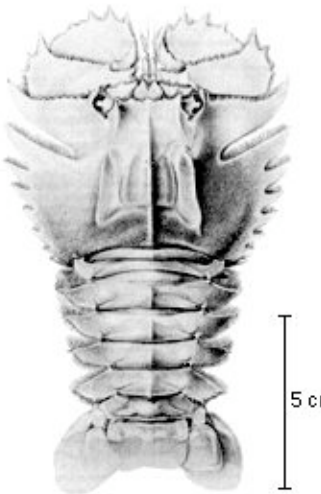


What kind of background knowledge will reveal this difference?

issues

AGROVOC NALT

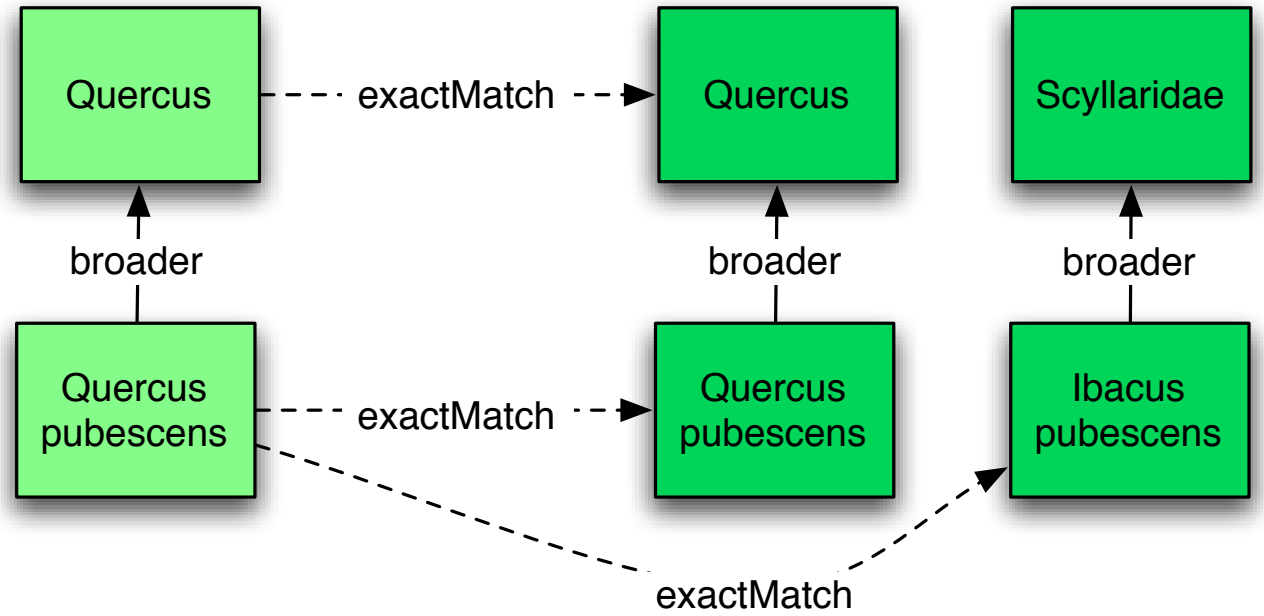


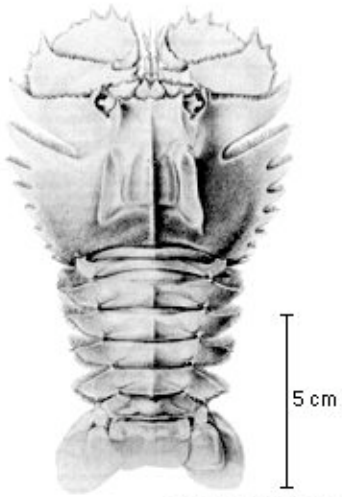


How can you define a generic rule about
pre- and postfixes in concept names?
(consider **sofsem:has_the_last_name** and **confious:last_name**)

issues

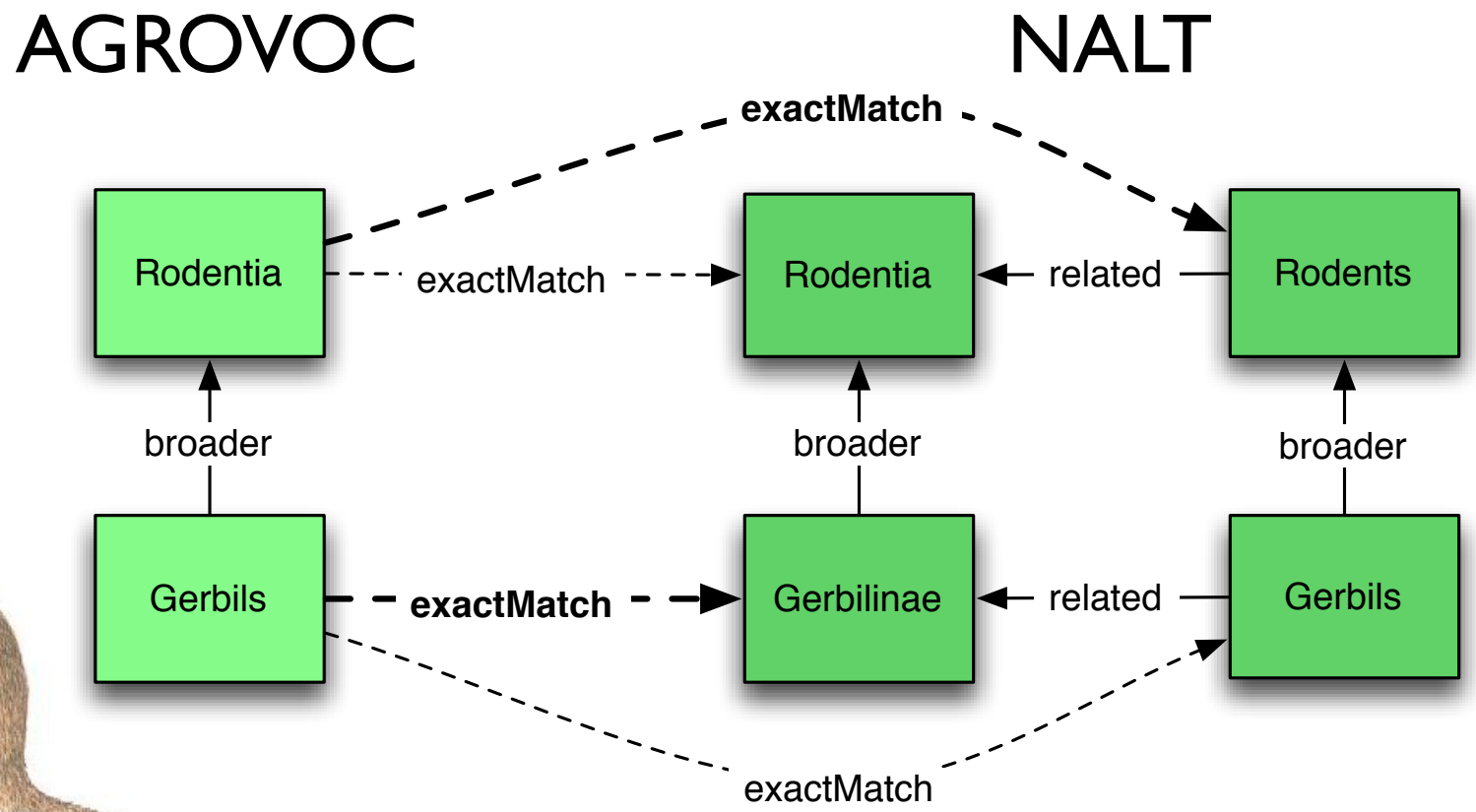
AGROVOC NALT





How can you define a generic rule about
pre- and postfixes in concept names?
(consider **sofsem:has_the_last_name** and **confious:last_name**)

issues



Are these exact matches?