

# OAEI 2007

## food & environment

Willem Robert van Hage

Free University Amsterdam  
and TNO Science & Industry

# food

## NALT & AGROVOC



AGROVOC  
Thesaurus

↑  
AGRIS  
CARIS



NAL  
Thesaurus

↑  
AGRICOLA  
AgNIC  
FSRIO  
NALDR



# environment

## GEMET, NALT & AGROVOC



GEMET  
Thesaurus

CDS



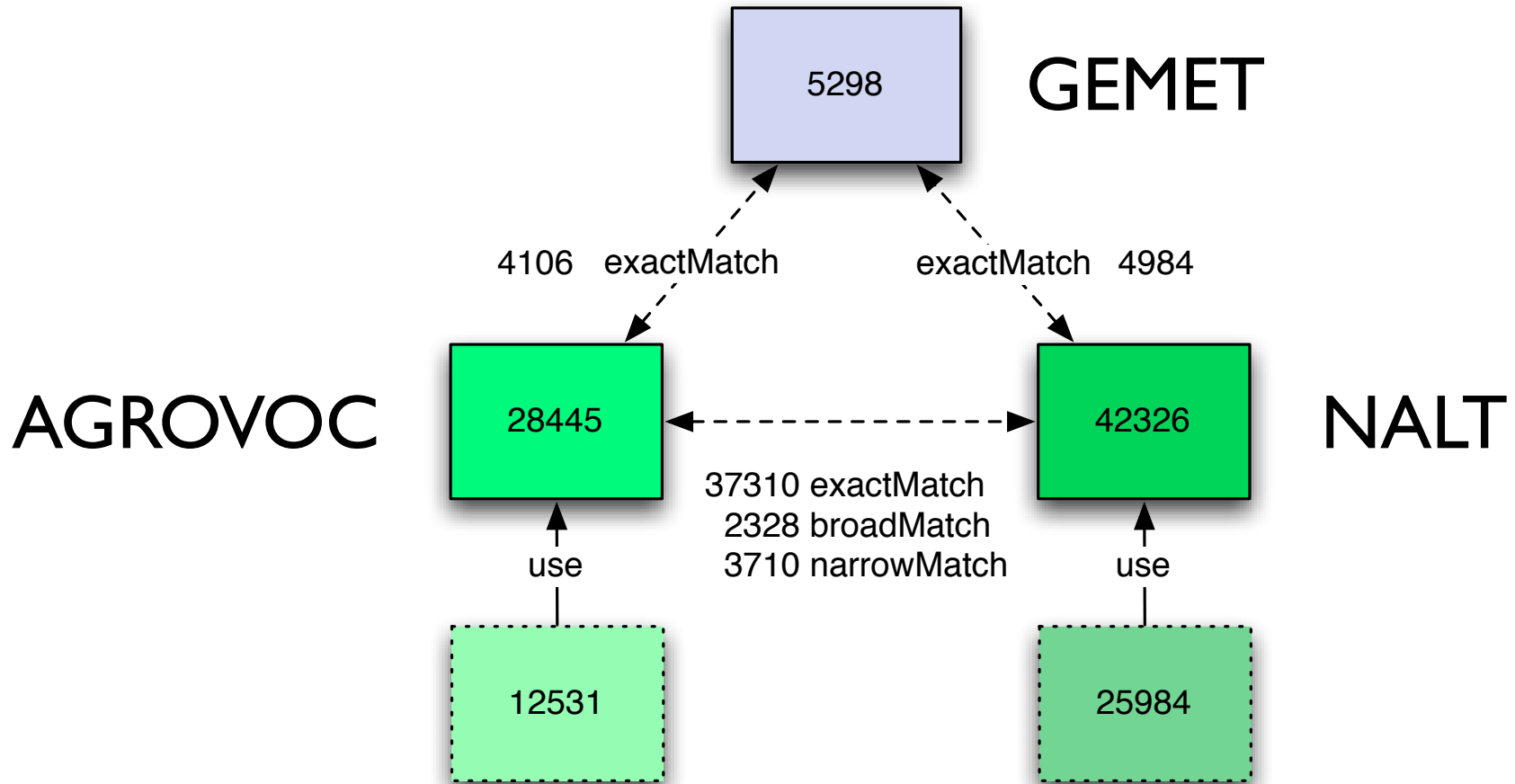
AGROVOC  
Thesaurus



NAL  
Thesaurus



# some numbers



# sample evaluation of precision

- domain experts evaluated 1,311 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
- 22,532 taxonomic mappings were evaluated semi-automatically

# assessment tool

34

1

Animals  
. Aphidoidea  
.. Adelgidae  
... **Adelges**  
alt: Chermes

exactMatch

Sternorrhyncha  
. Aphidoidea  
.. Adelgidae  
... **Adelges**  
alt: Chermes  
... Adelges tsugae

2

... Adelges

exactMatch

... Adelges

3

unknown

# sample evaluation of recall

- we estimate recall based on a set of 826 mappings from various domains
  - NALT-AGROVOC
    - “animal health”, “geography of europe”, “geography below country level” (e.g. provinces), “rodents”, and “oak trees”
  - GEMET-AGROVOC, GEMET-NALT
    - “agriculture” (also forestry, aquaculture, etc.), “geography / geology” (locations, natural resources, etc.)
- the samples were made by manually creating a complete mapping for these subjects

# manual mapping tool

AIDA Thesaurus Browser

Thesaurus Browser & Mapping Editor

Add a  mapping between these concepts.

Search:

[show concept schemes](#)

## Maize

<http://www.fao.org/aos/agrovoc>

Corn (maize)

**in concept scheme**

- [General \(not included in previous categories\)](#) (39777)

**broader terms**

- [Cereals](#) (20)

**narrower terms**

- [Dent maize](#) (0)
- [Flint maize](#) (0)
- [Popcorn](#) (0)
- [Soft maize](#) (0)
- [Sweet corn](#) (0)
- [Waxy maize](#) (0)

**related terms**

- [Corn flour](#) (0)
- [Corn starch](#) (0)
- [Zea mays](#) (0)

Search:  [close this view](#)

[show concept schemes](#)

## corn

<http://agclass.nal.usda.gov/nalt/2006.xml>

feed corn  
field corn  
maize

**note** Use for corn as a grain; For corn as a vegetable USE sweetcorn; For the corn plant USE Zea mays.

**broader terms**

- [grains](#) (33)

**narrower terms**

- [dent corn](#) (0)
- [flint corn](#) (0)
- [popcorn](#) (0)
- [soft corn](#) (0)
- [waxy corn](#) (0)

**related terms**

- [corn products](#) (15)
- [nixtamalization](#) (0)
- [sweetcorn](#) (0)
- [Zea mays](#) (4)

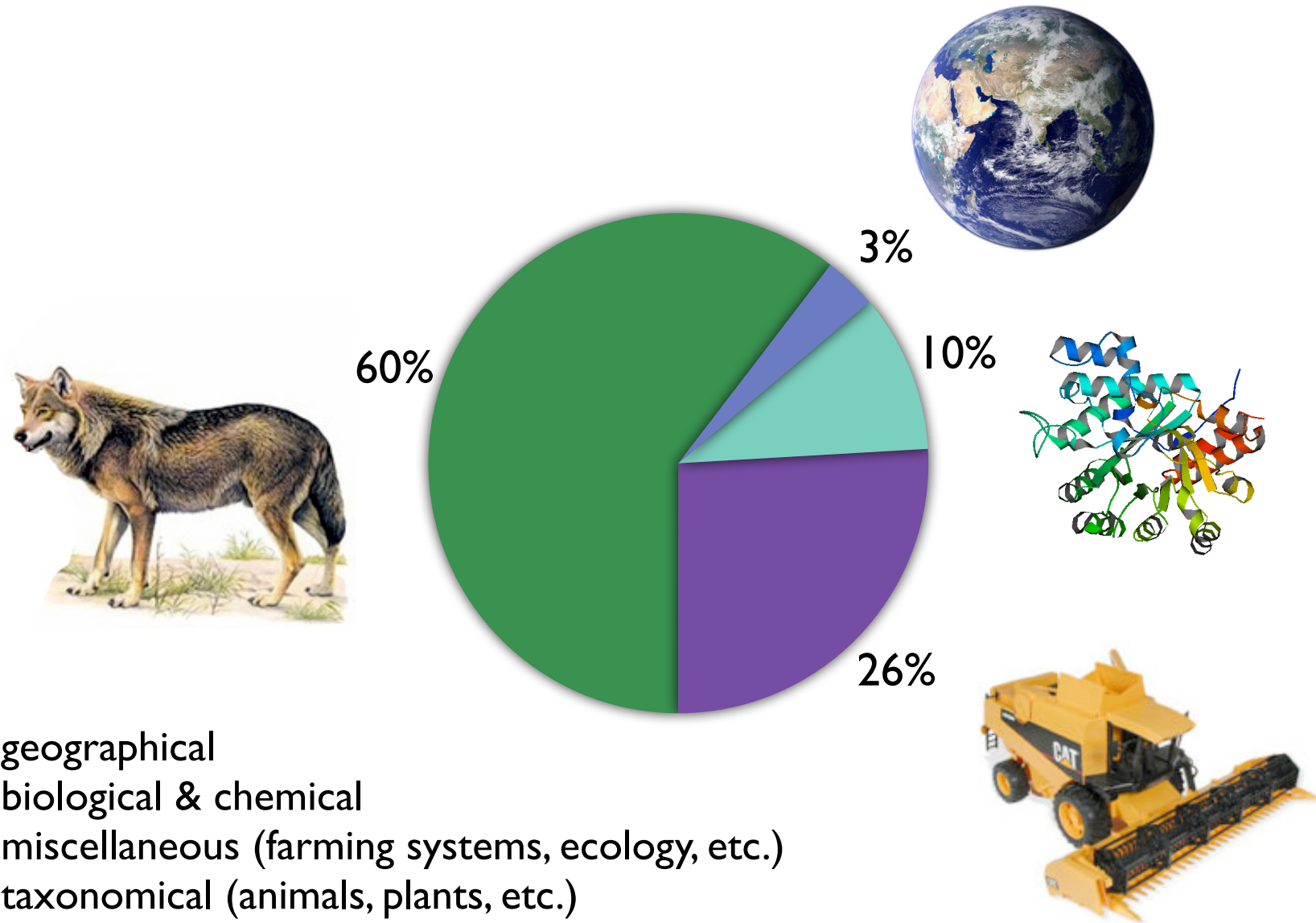


food

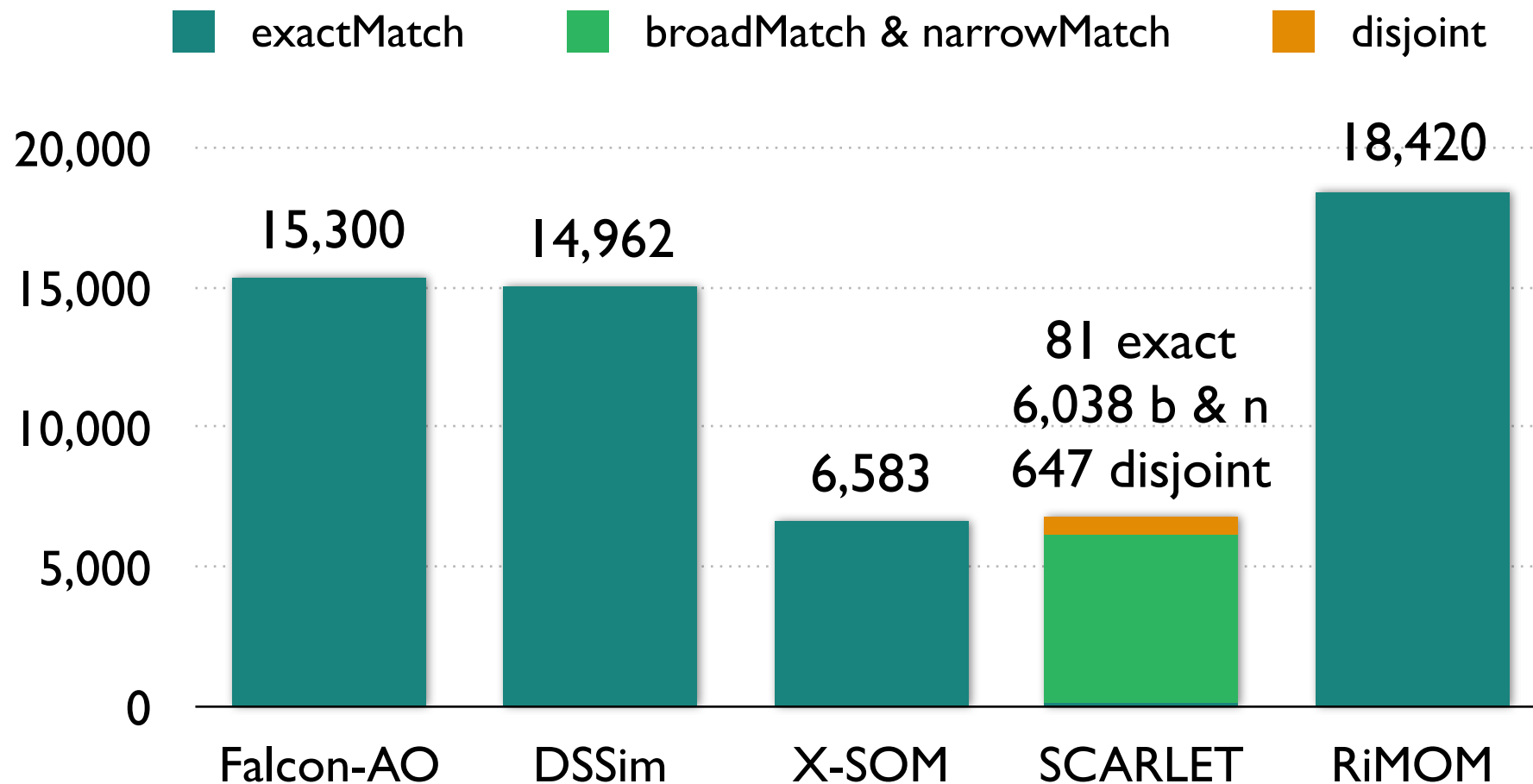
**NALT-AGROVOC**

# topics in the results

NALT-AGROVOC

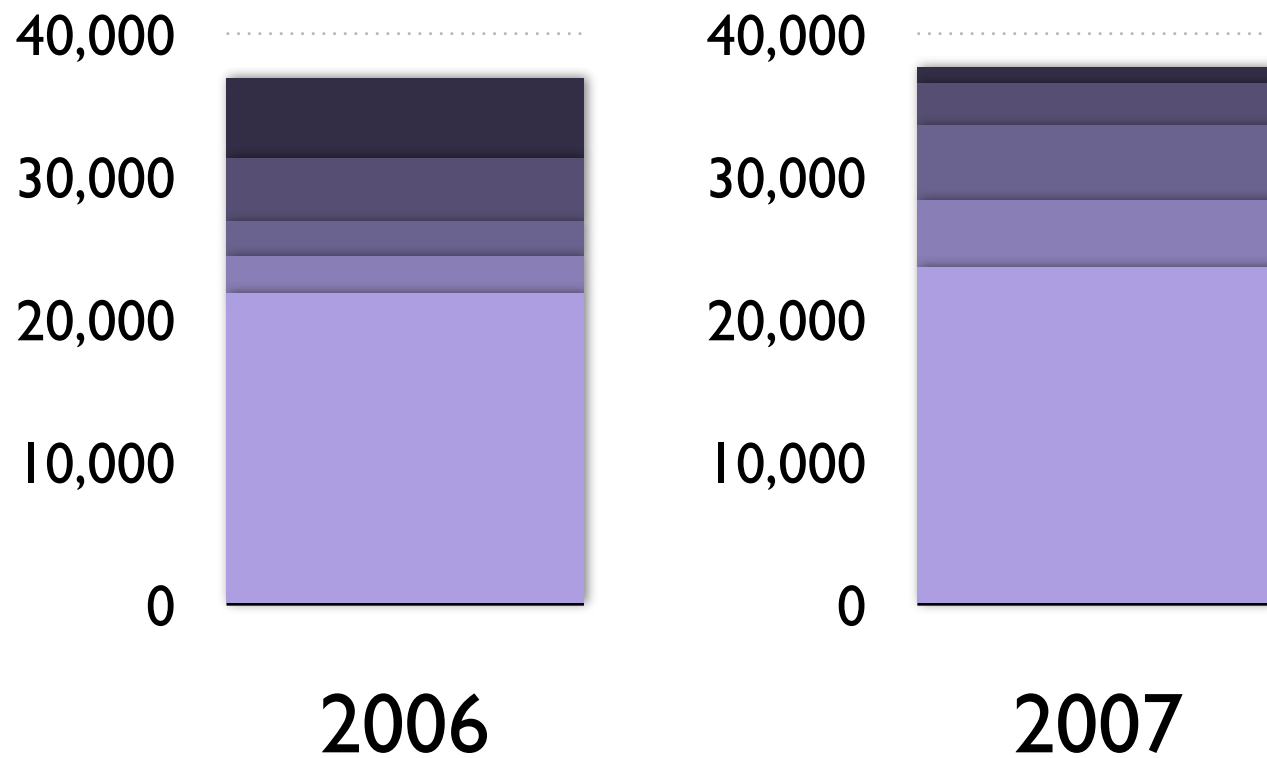


# number of results

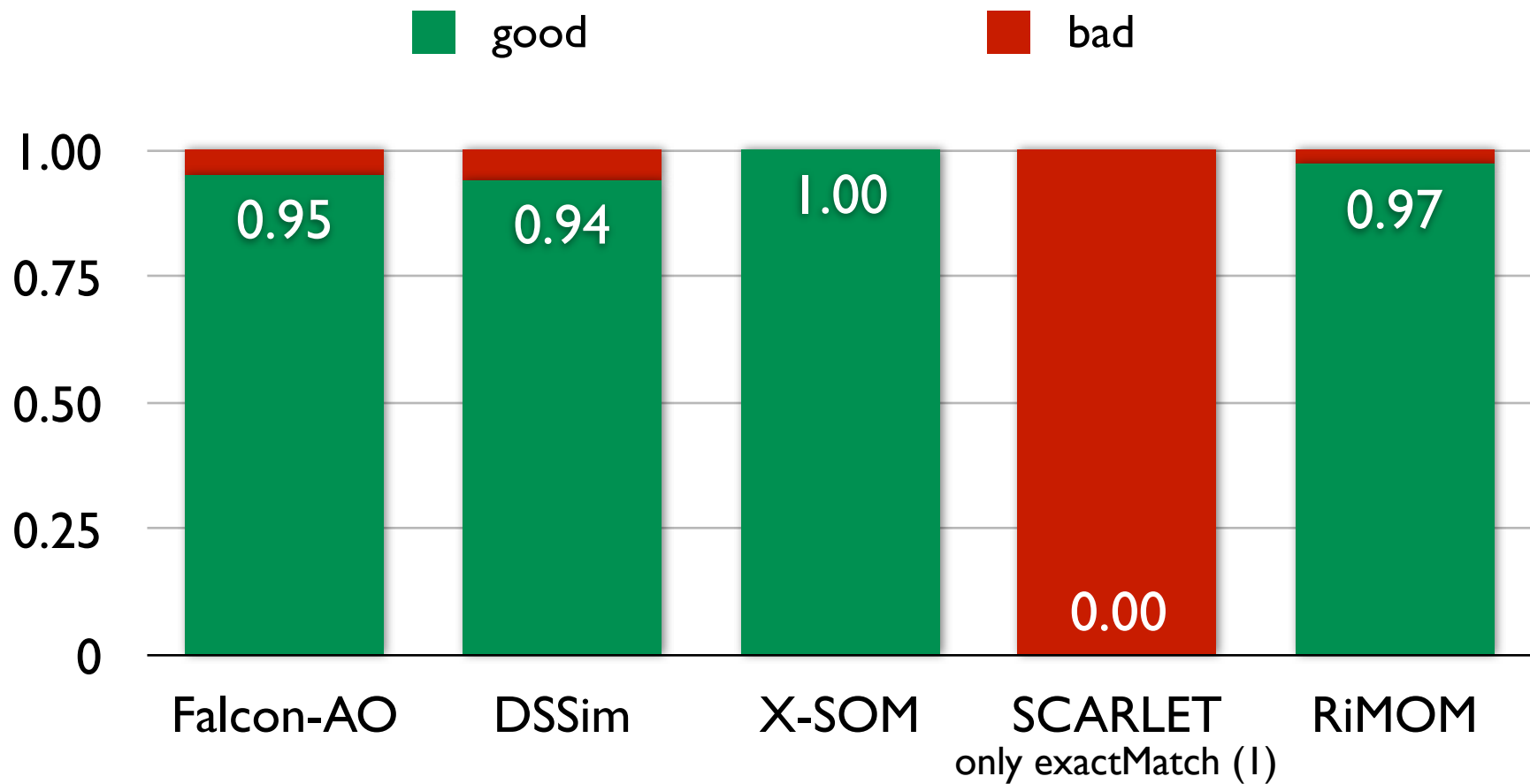


# result overlap

submitted by # participants

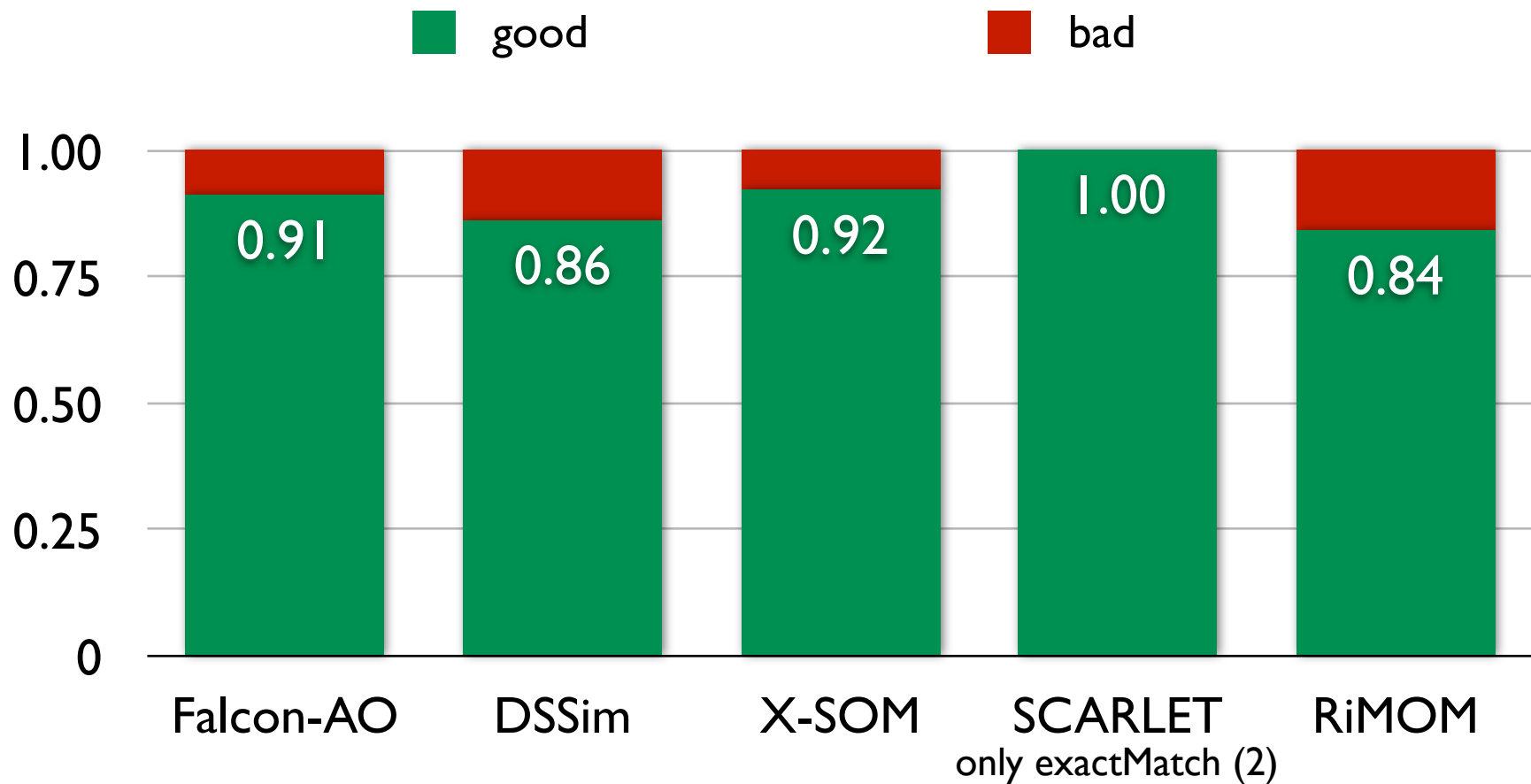


# precision of geographical mappings (3%)



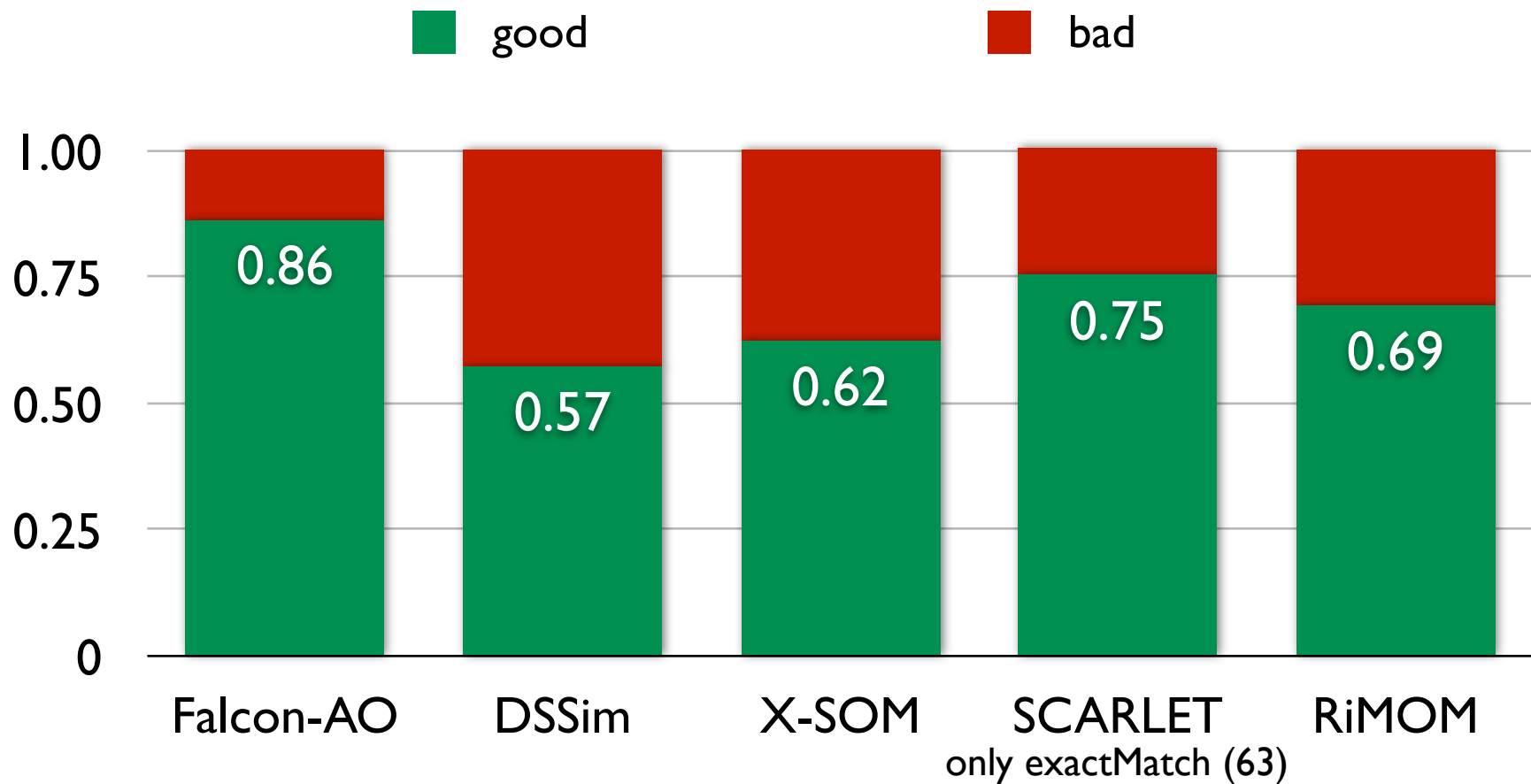
both political and geographical areas, countries, rivers, etc.

# precision of biological & chemical mappings (10%)



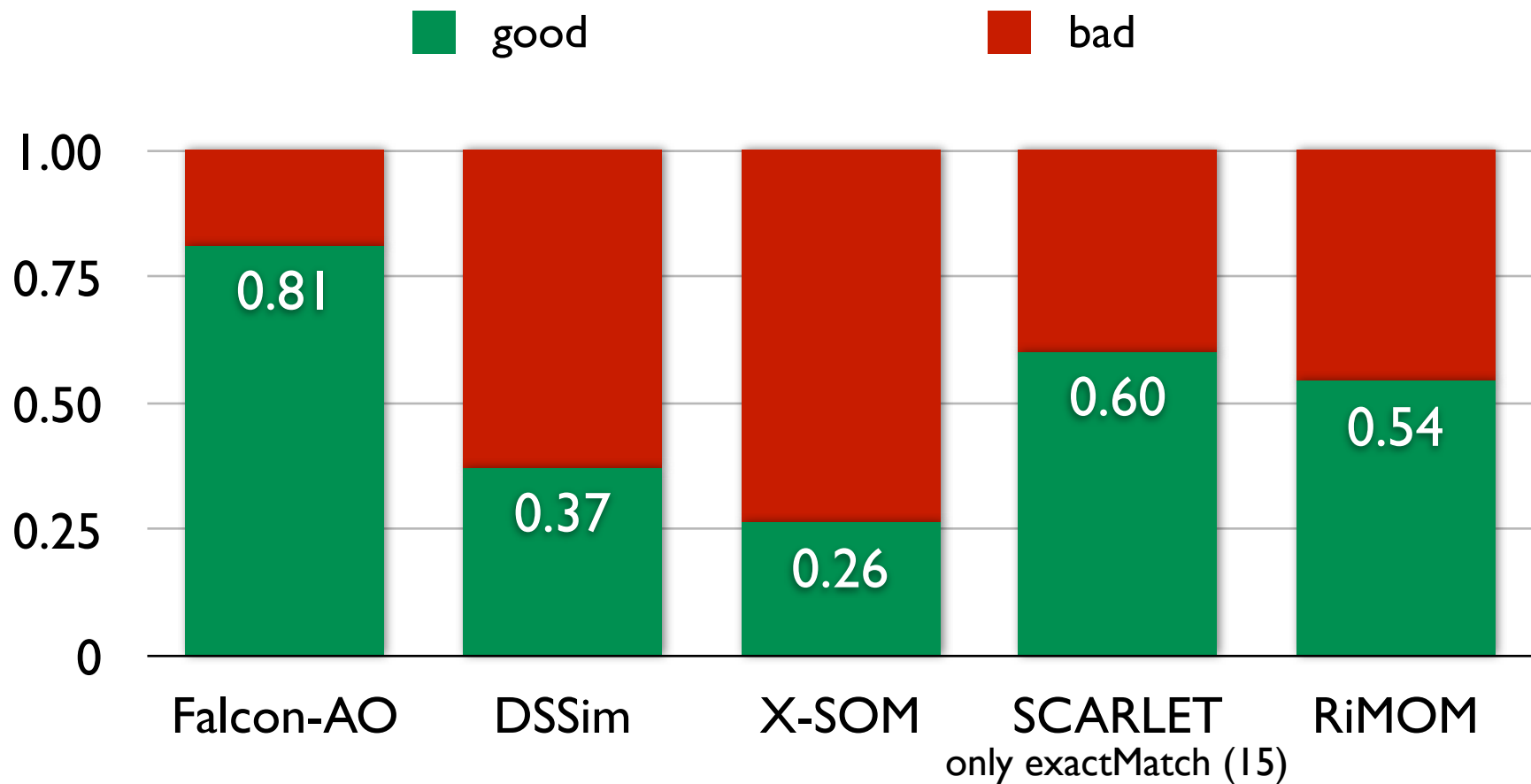
very specific terms, many spelling variants  
different terms may mean the same thing

# precision of miscellaneous mappings (26%)



diverse topics, similar terms usually have similar meaning

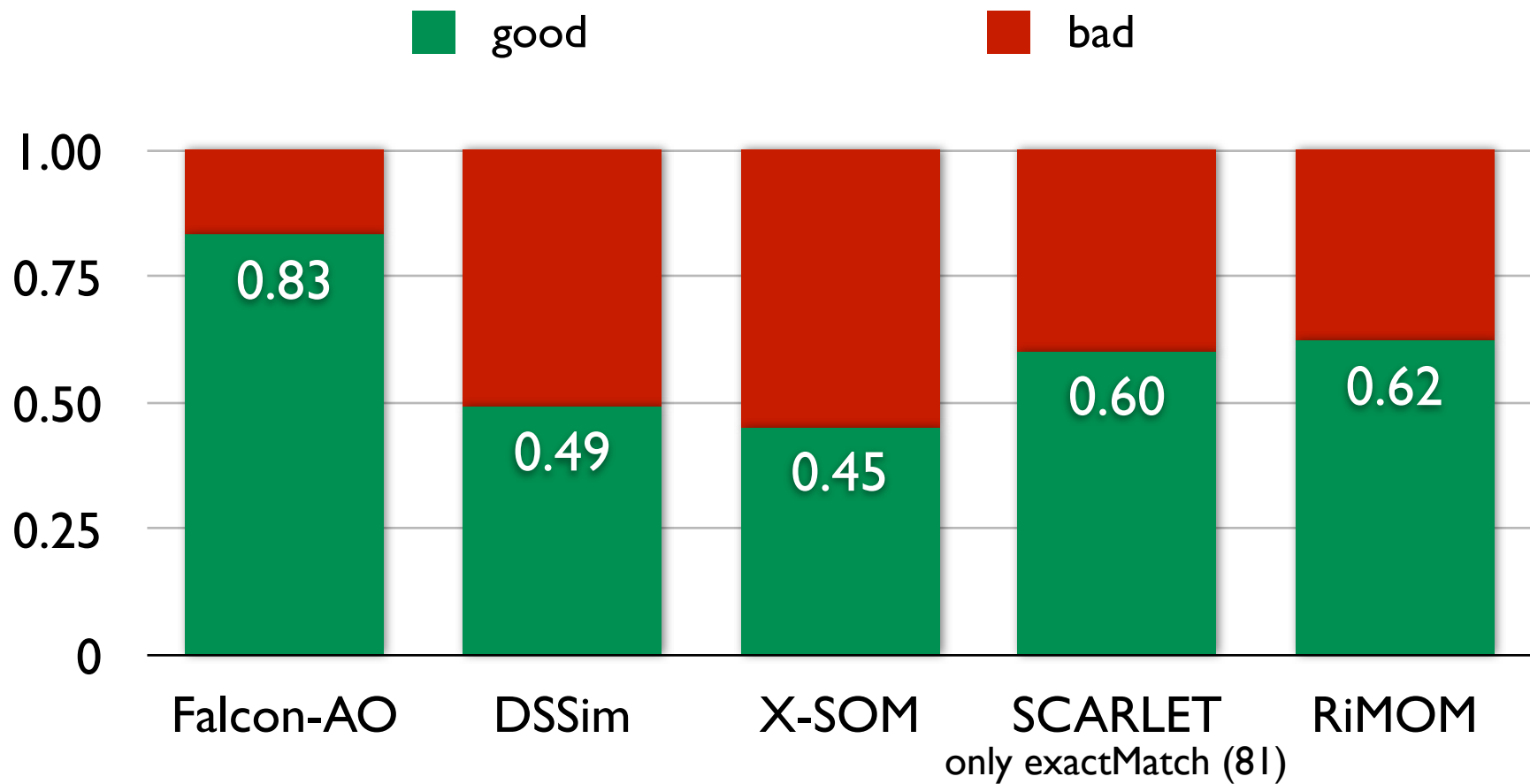
# precision of taxonomical mappings (60%)



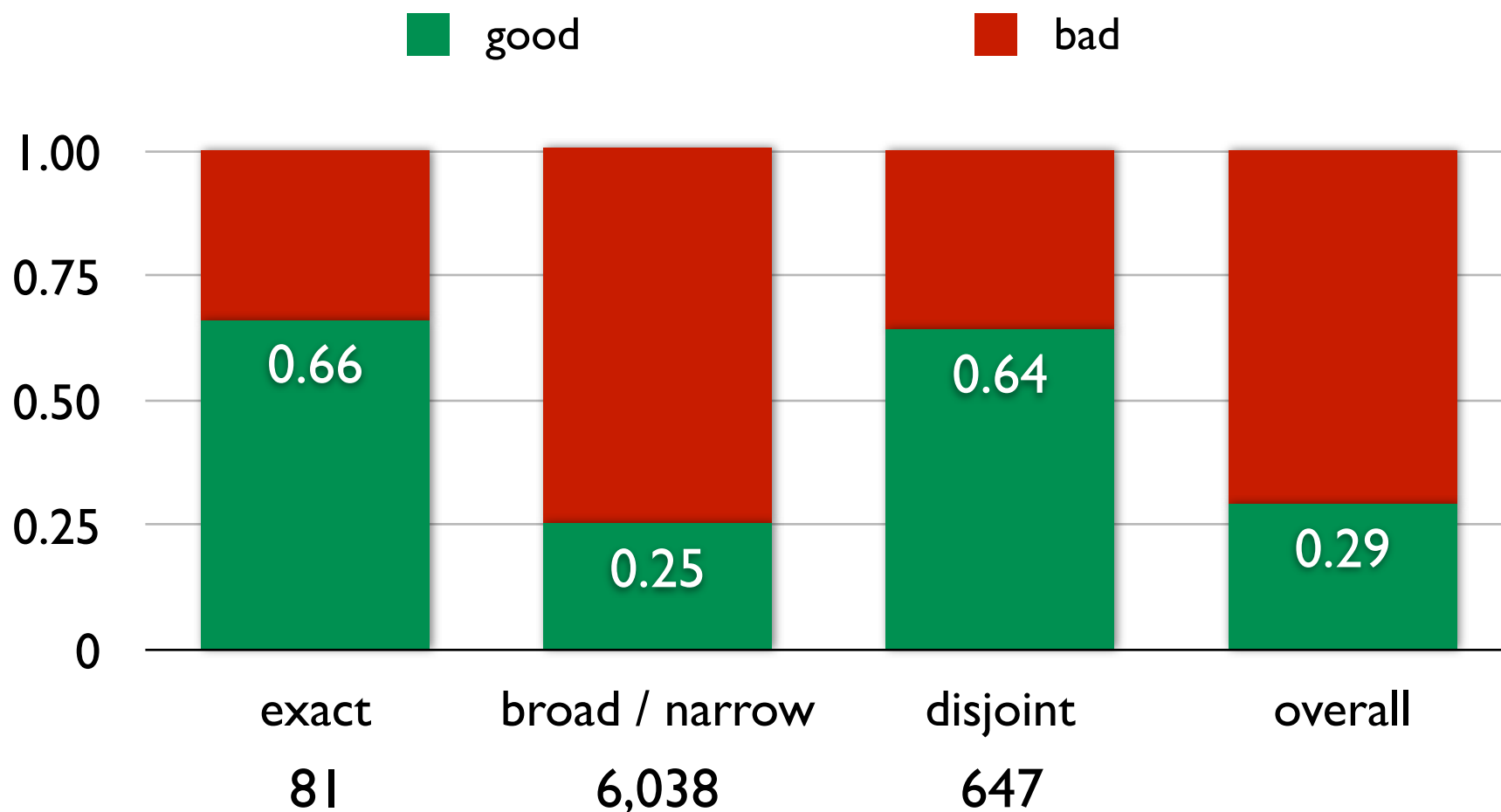
many similar terms with different meaning  
e.g. Canis Familiaris, Canis Lupus



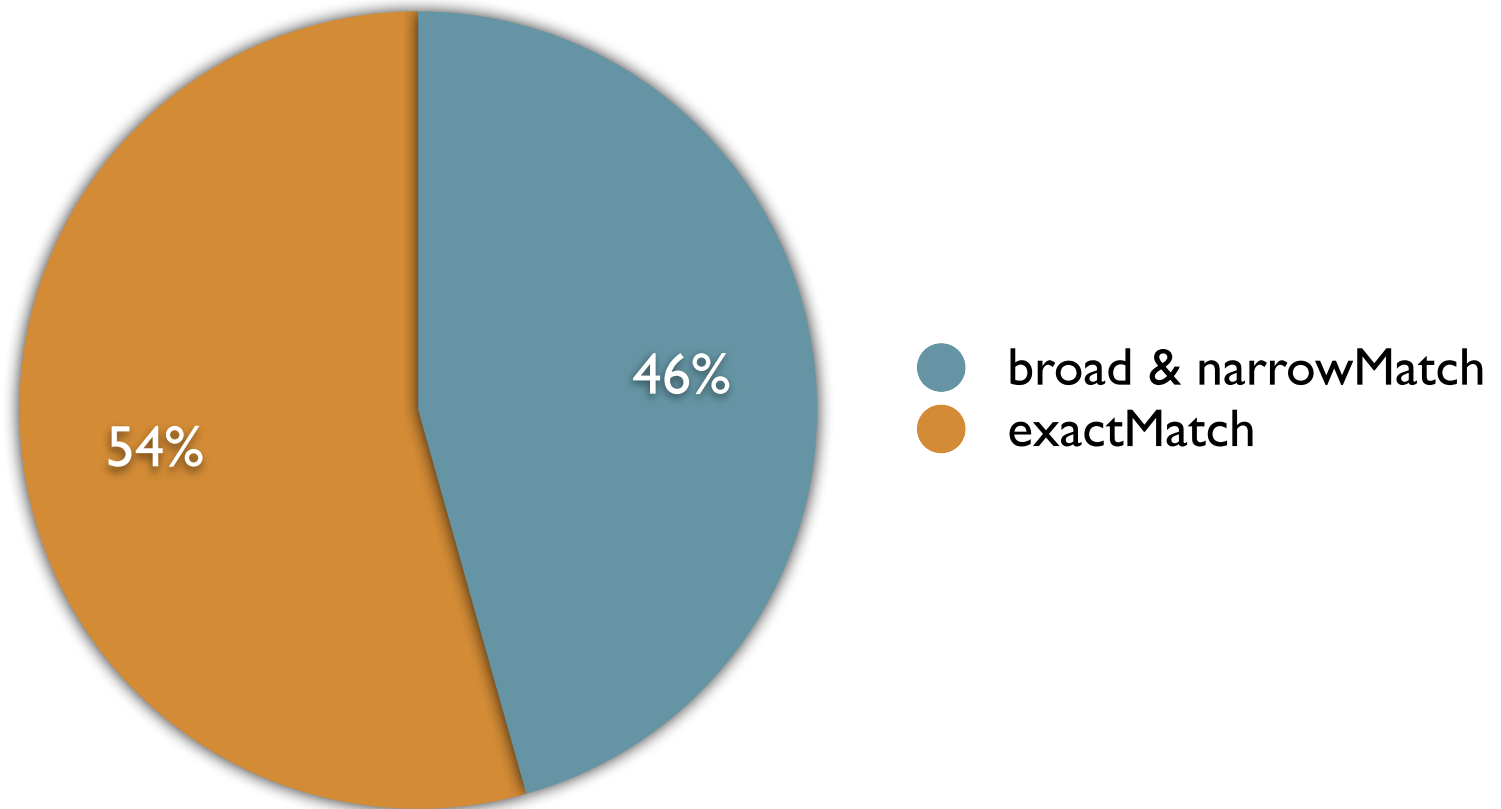
# overall precision (100%)



# SCARLET



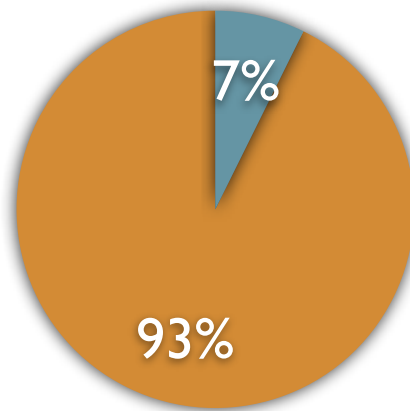
# gold standard relations



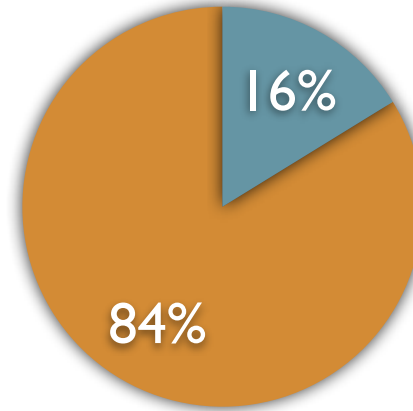
all topics

# gold standard relations

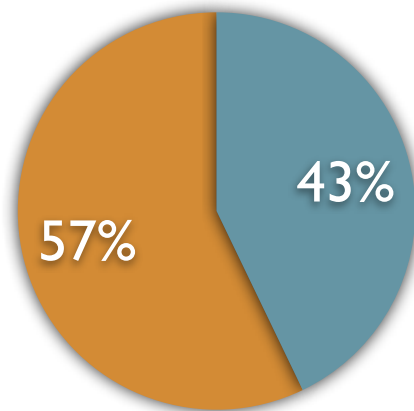
- broad & narrowMatch
- exactMatch



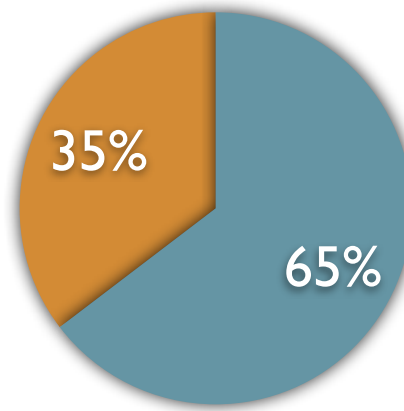
Europe



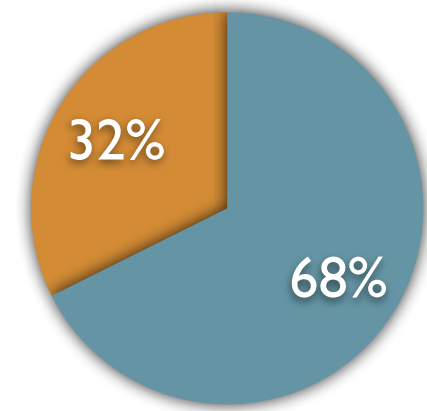
Oak trees



Animal health

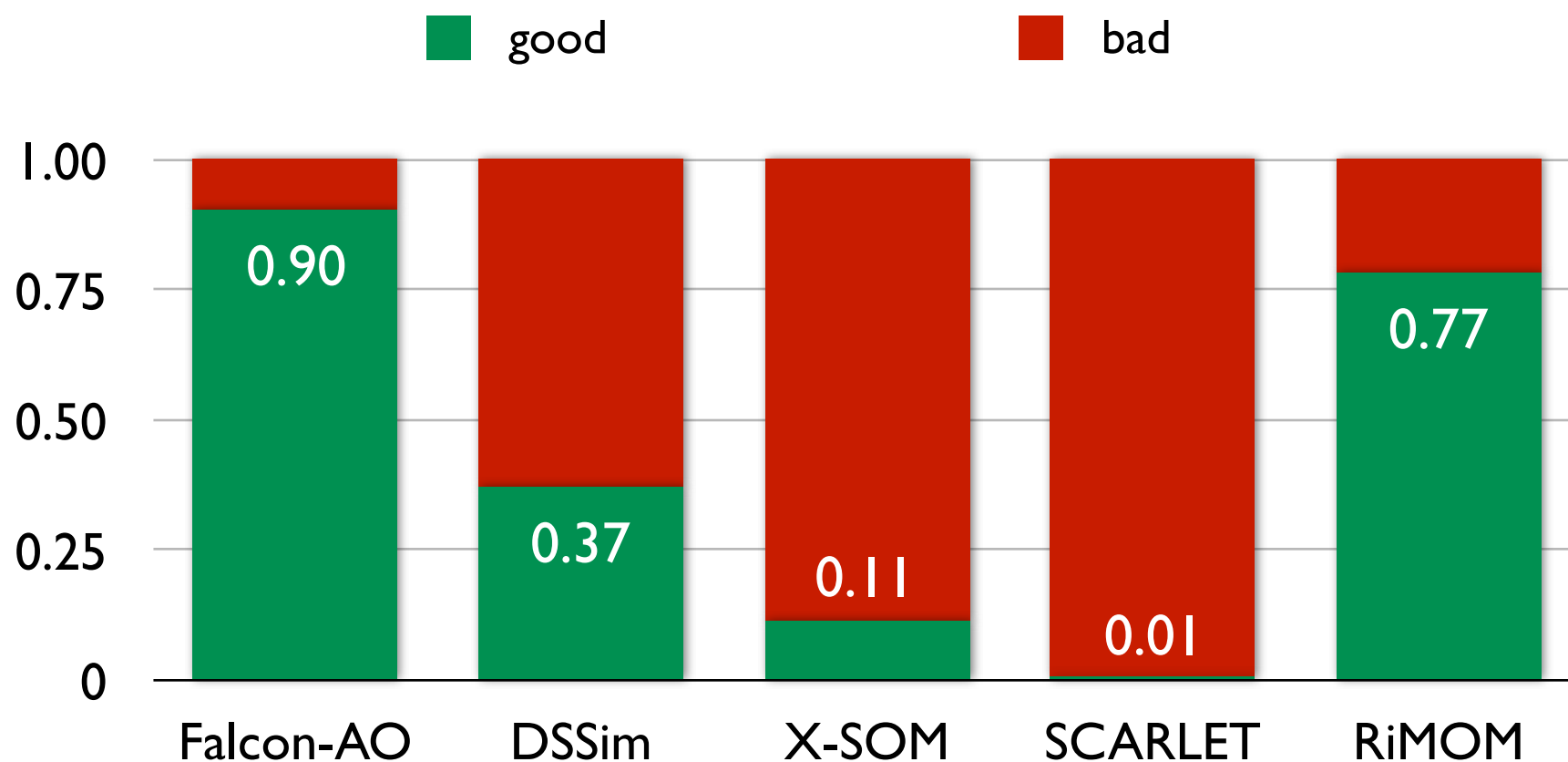


Topography  
below country level



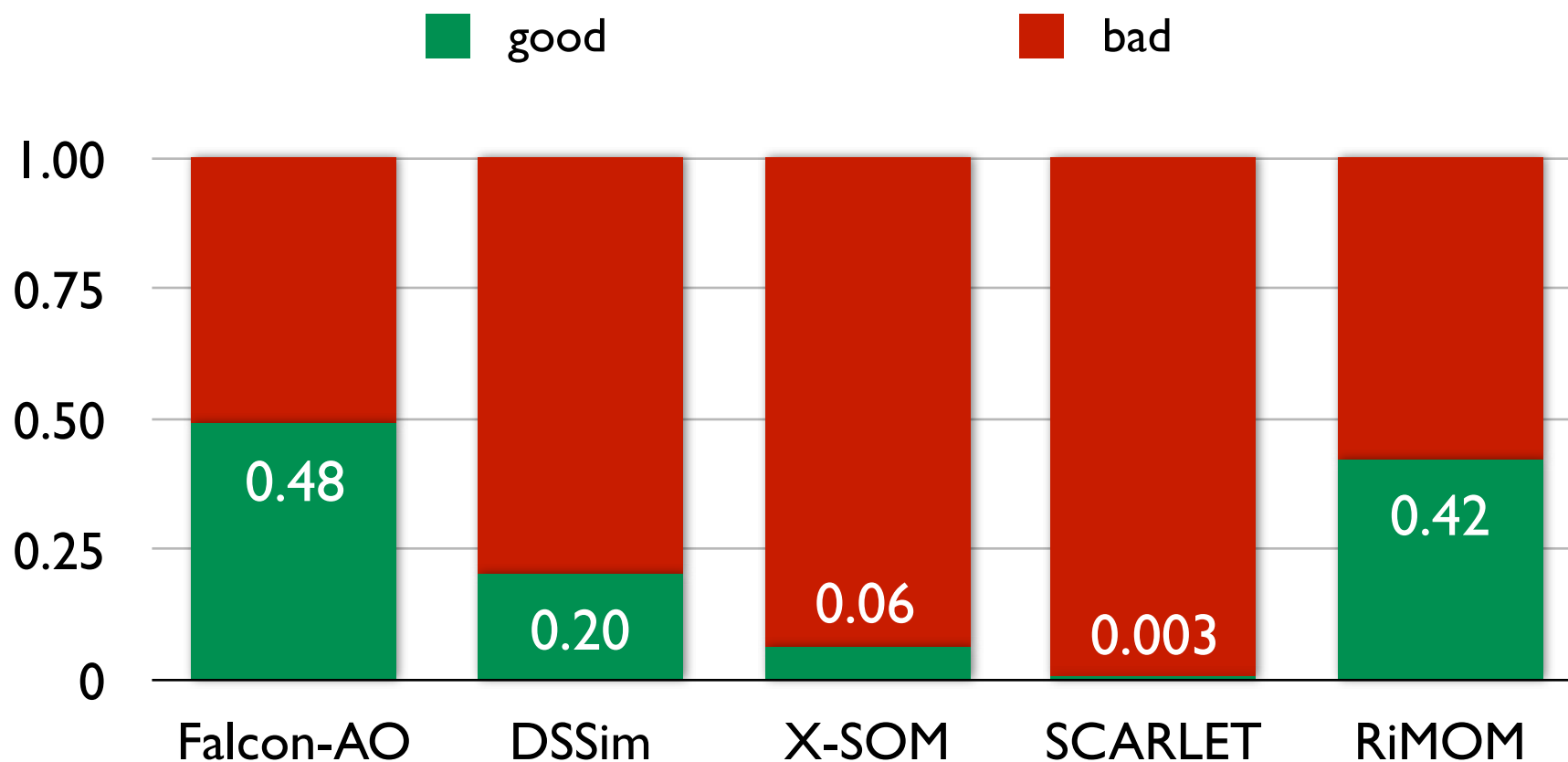
Rodents

# recall of only exactMatch



both political and geographical areas, countries, rivers, etc.

# recall of exact, broad & narrowMatch



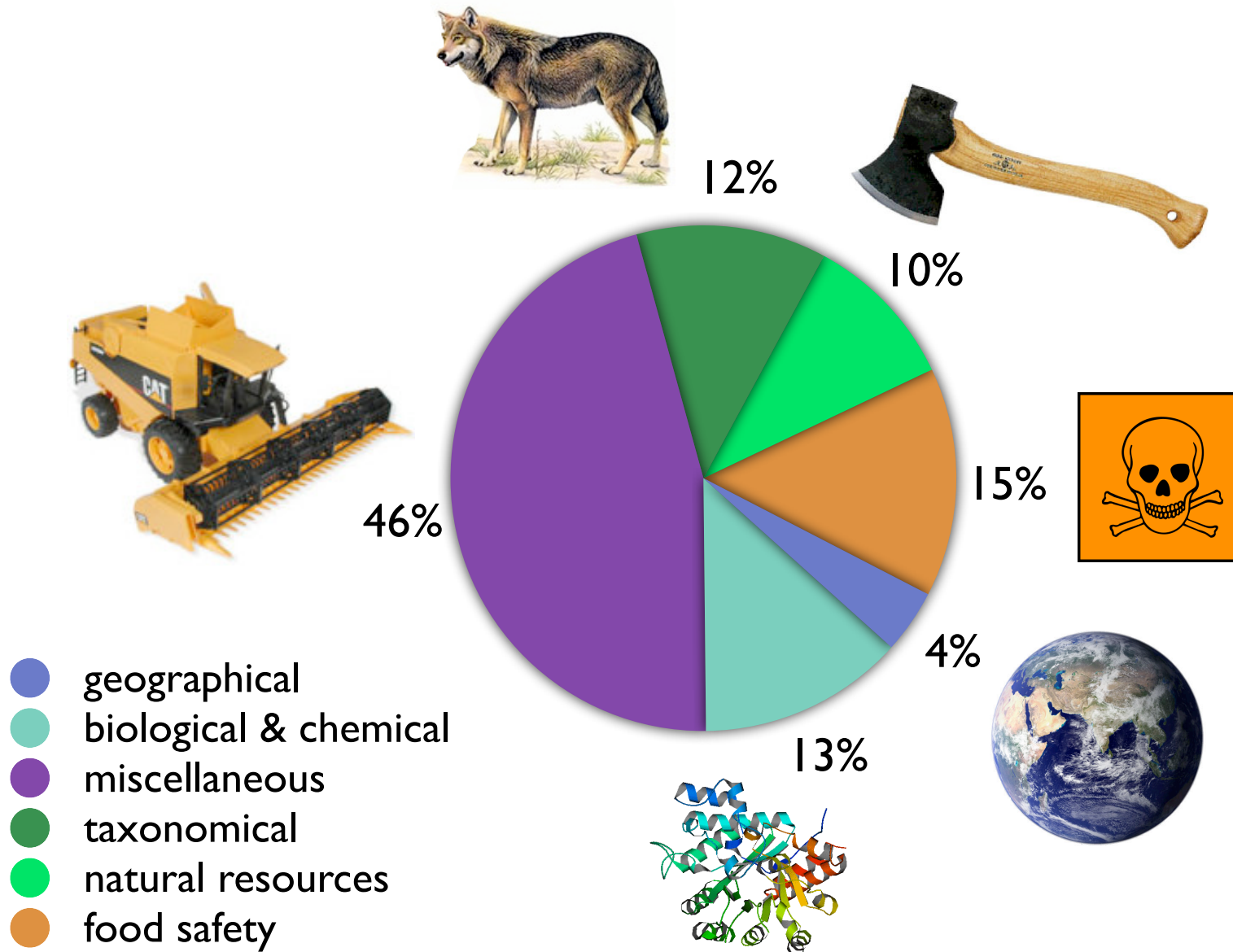
both political and geographical areas, countries, rivers, etc.

environment

**GEMET-AGROVOC  
& GEMET-NALT**

# topics in the results

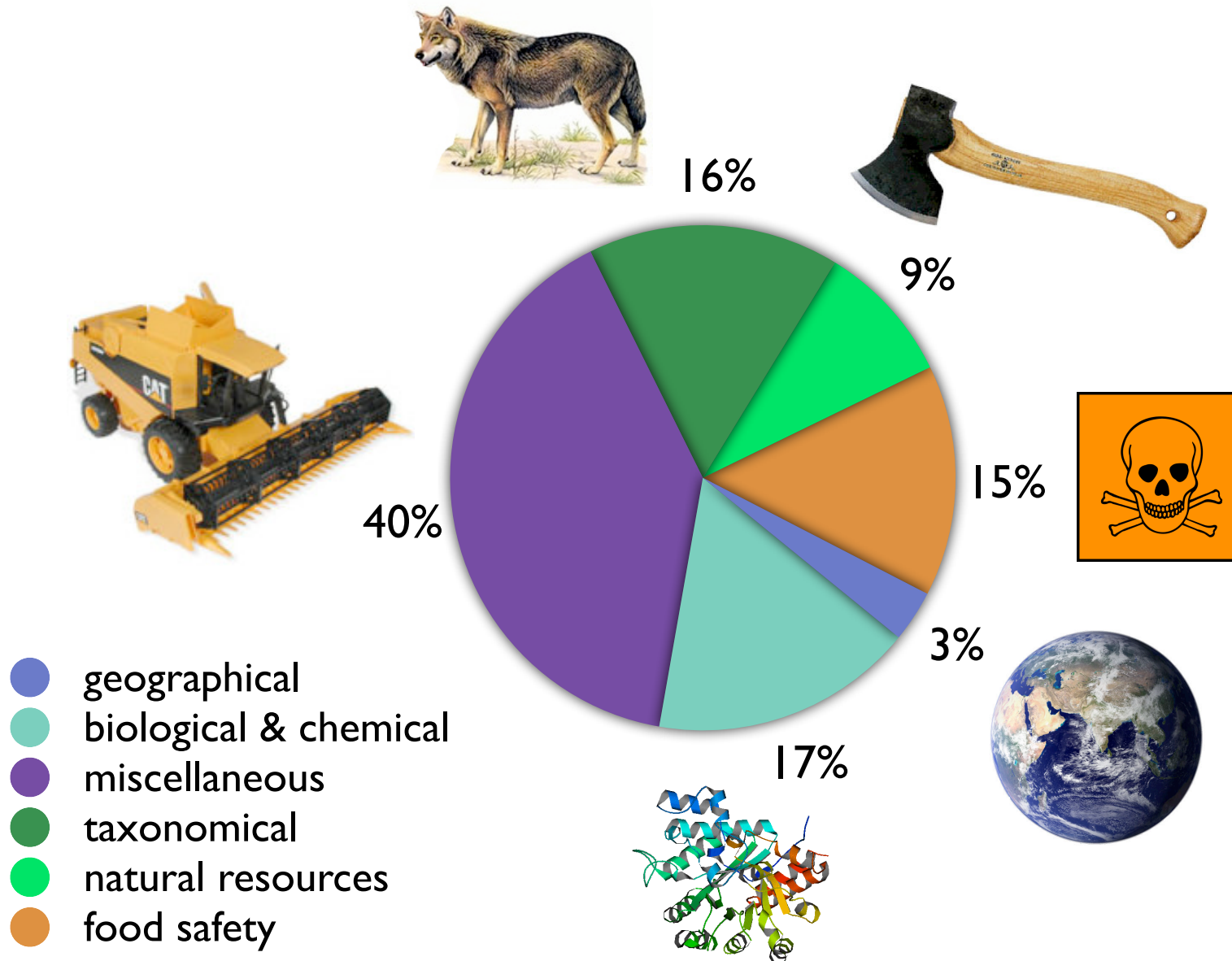
## GEMET-AGROVOC





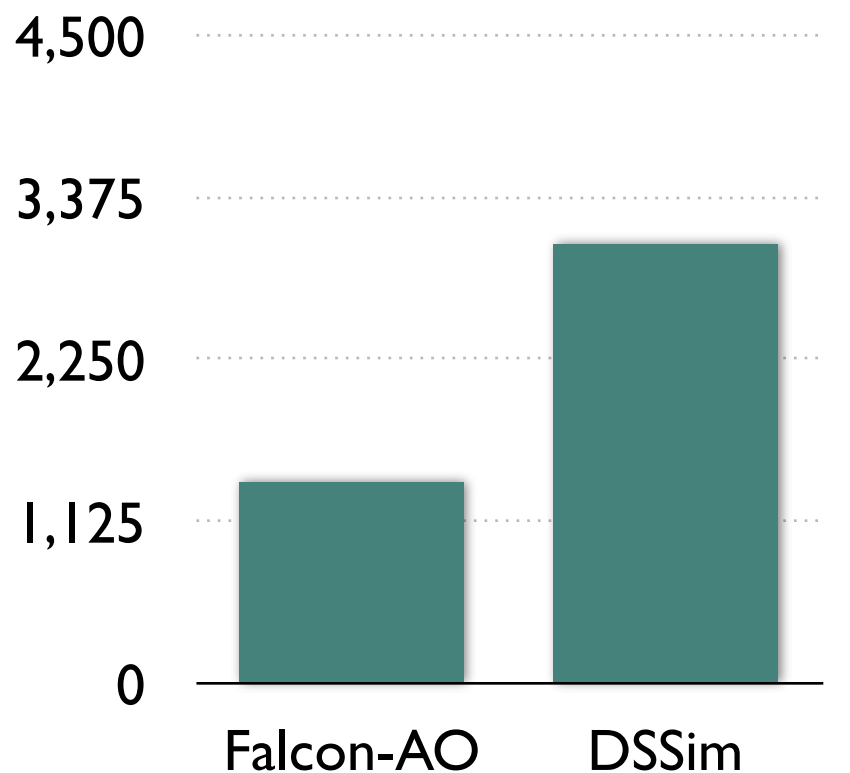
# topics in the results

GEMET-NALT

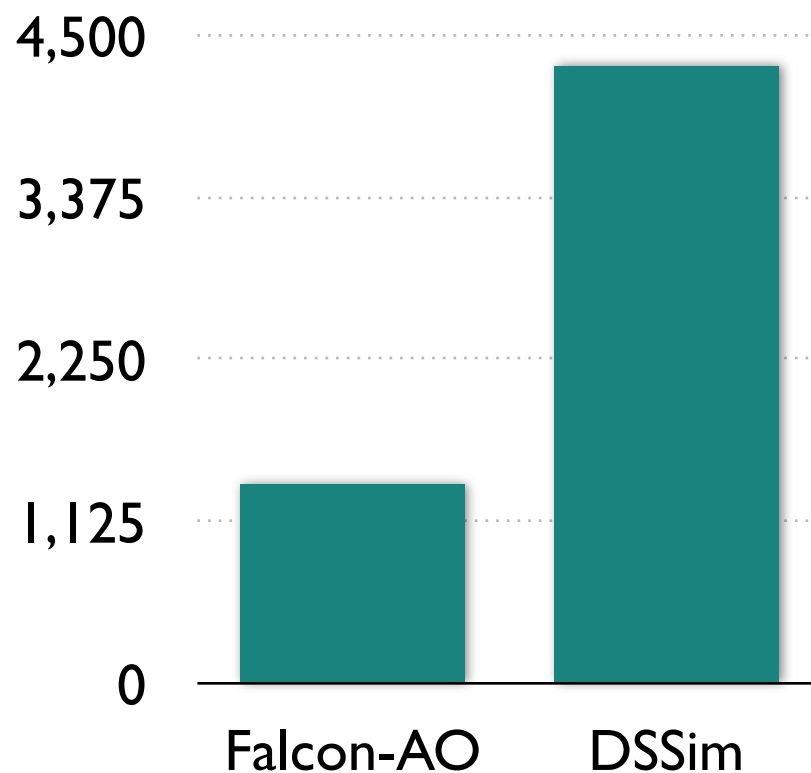


# number of results

exactMatch    broadMatch & narrowMatch    disjoint



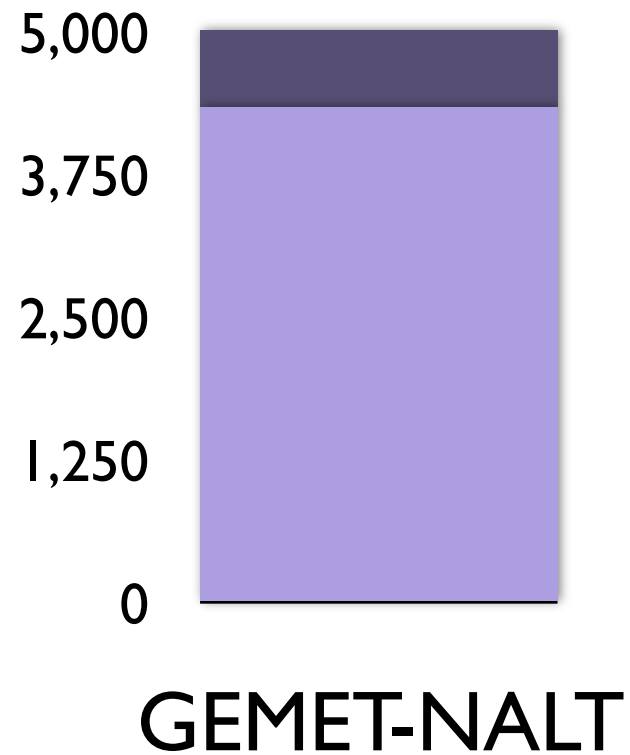
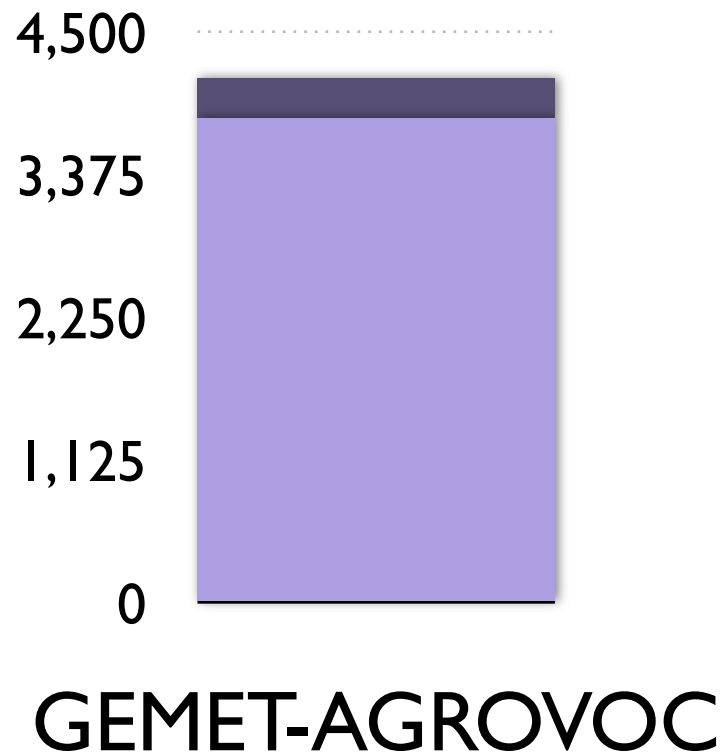
GEMET-AGROVOC



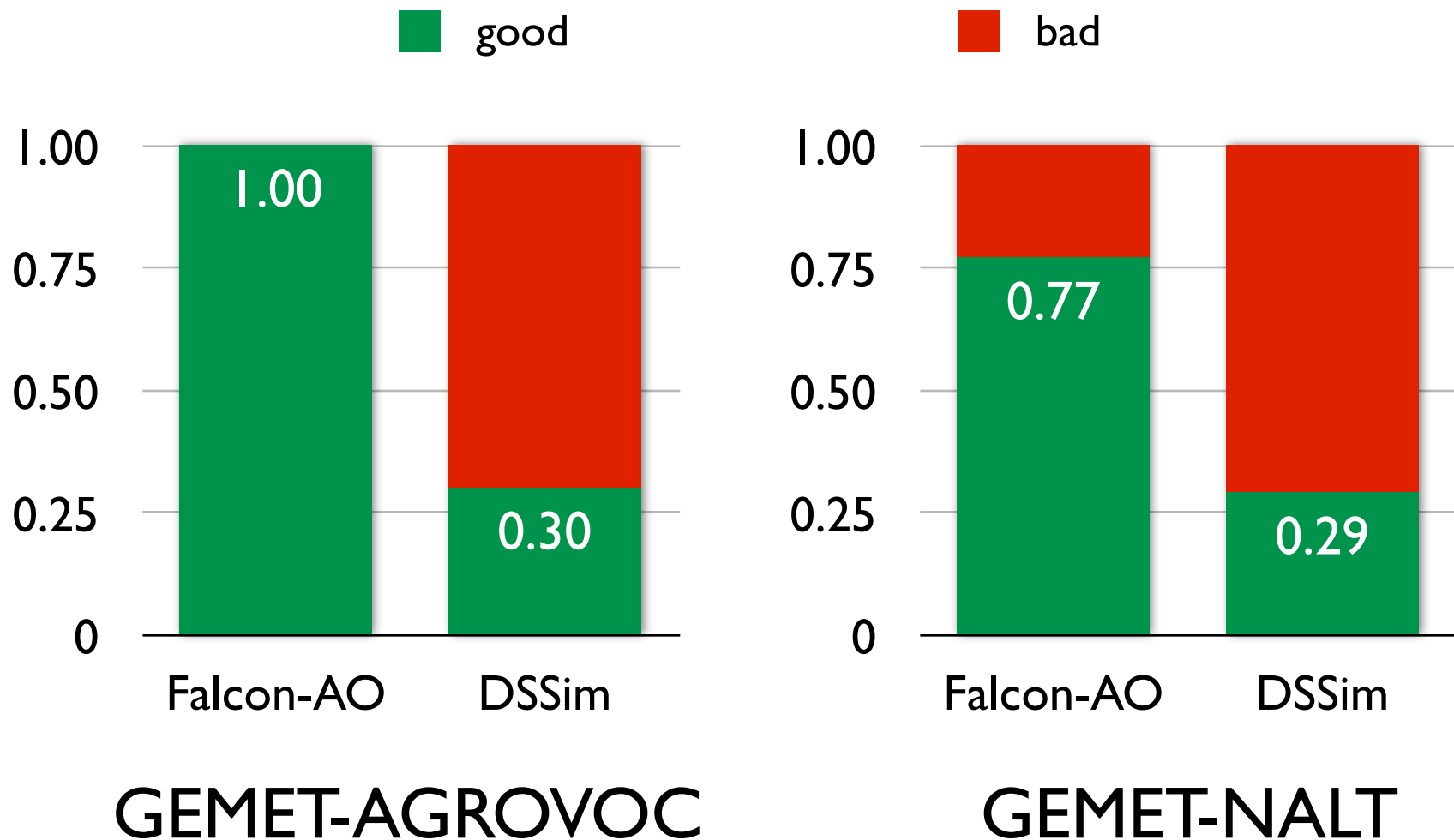
GEMET-NALT

# result overlap

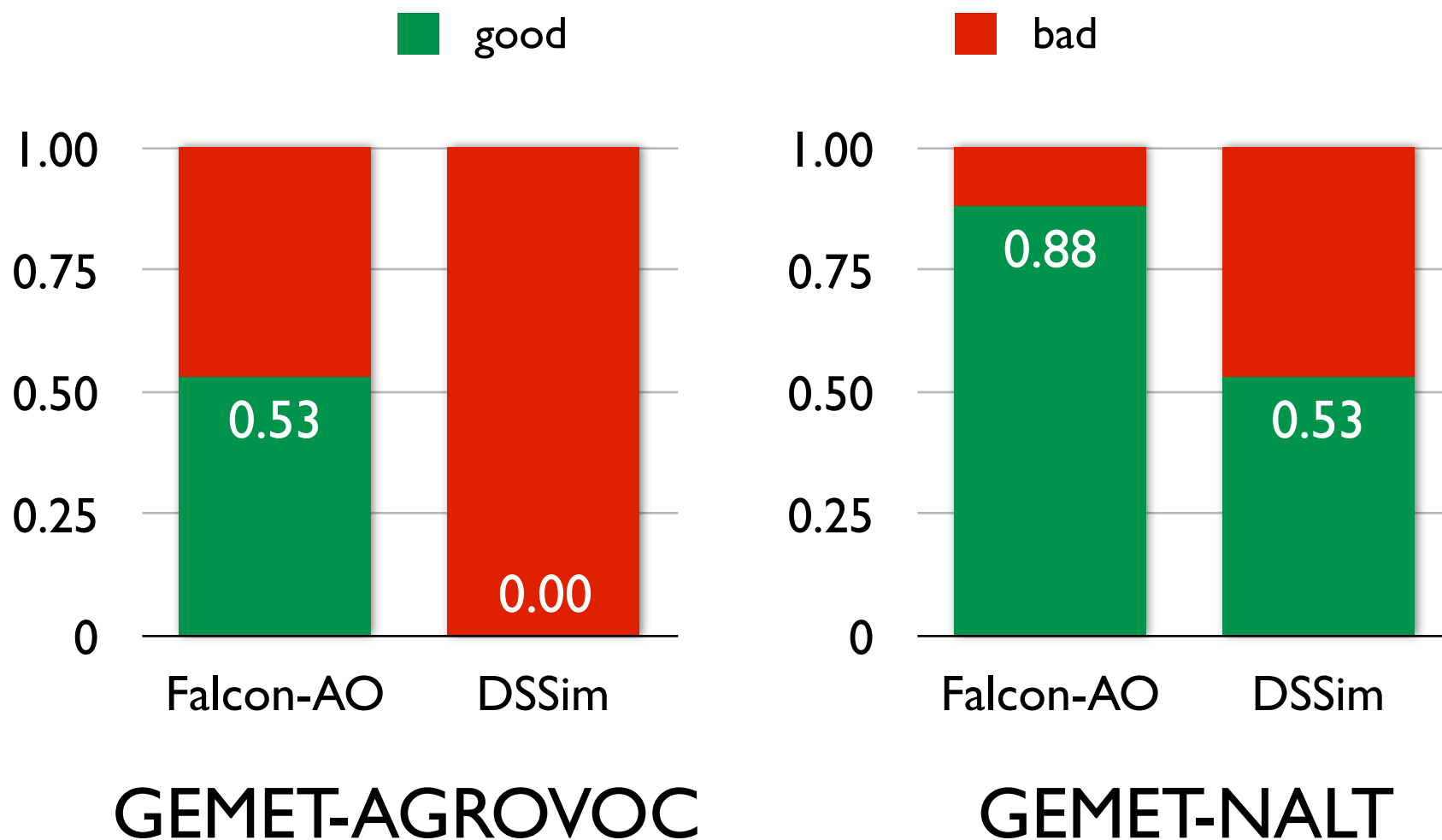
submitted by # participants



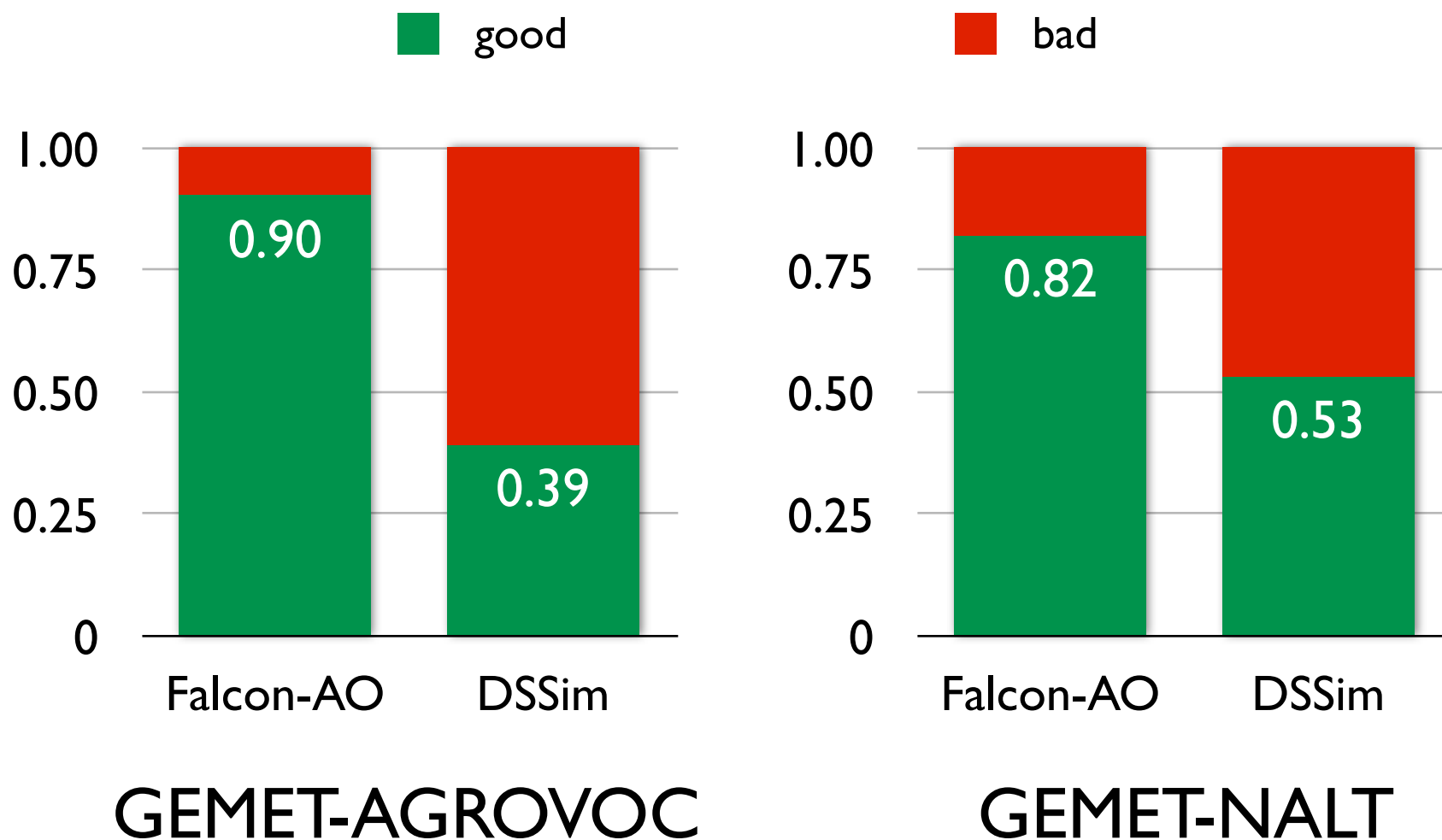
# precision of geographical mappings (4%, 3%)



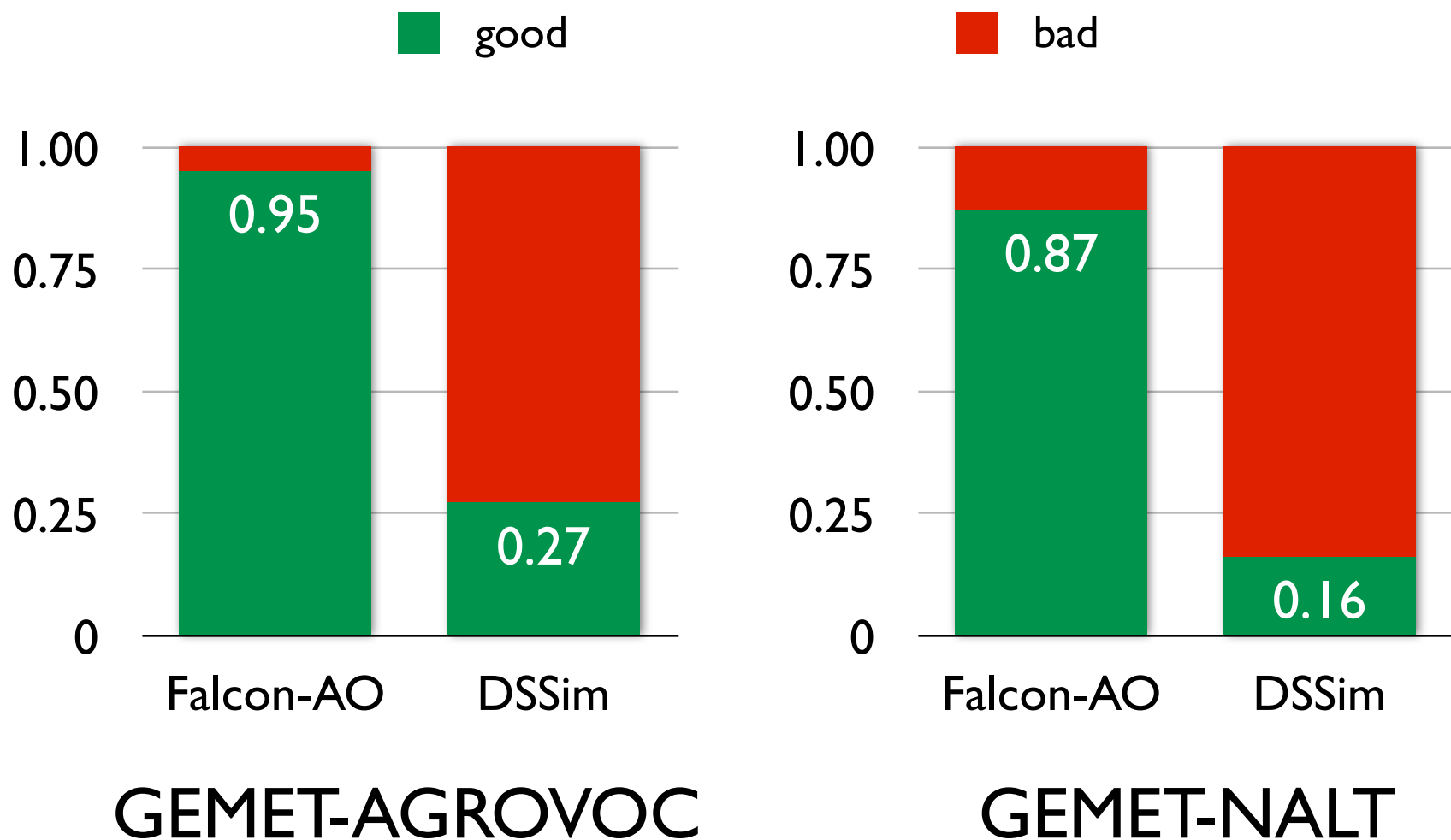
# precision of biological & chemical mappings (13%, 17%)



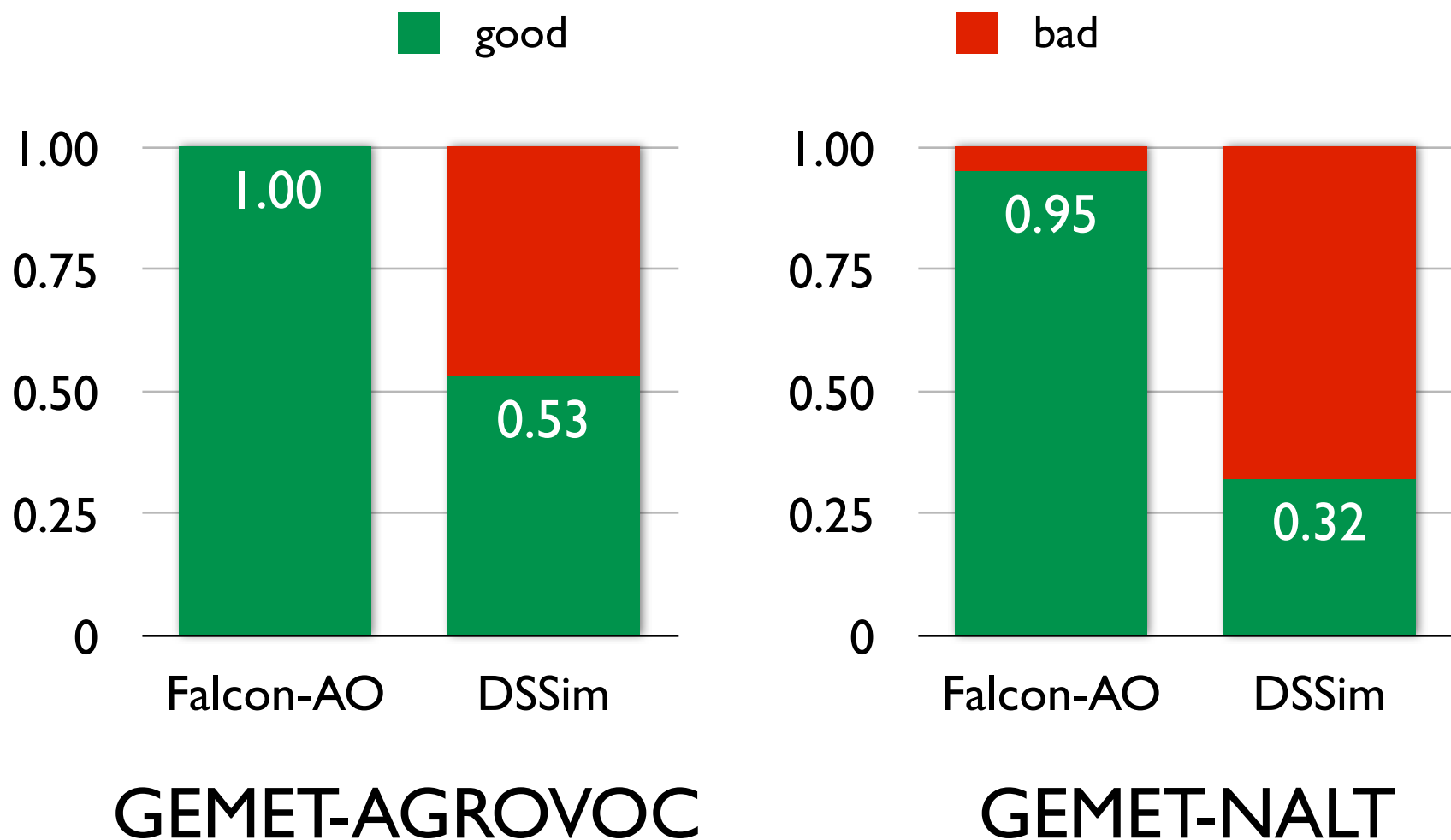
# precision of miscellaneous mappings (46%, 40%)



# precision of taxonomical mappings (12%, 16%)

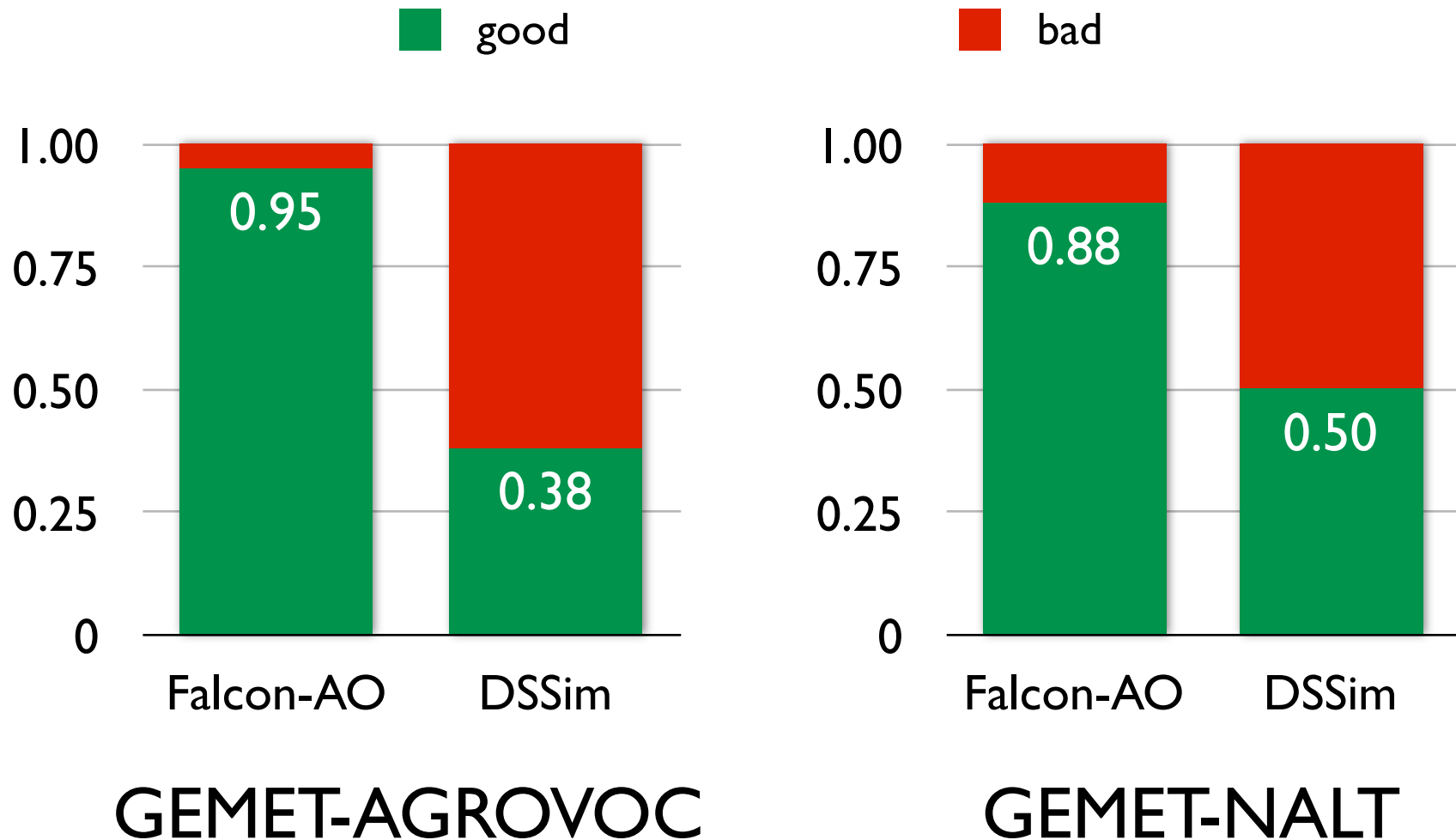


# precision of natural resources mappings (10%, 9%)

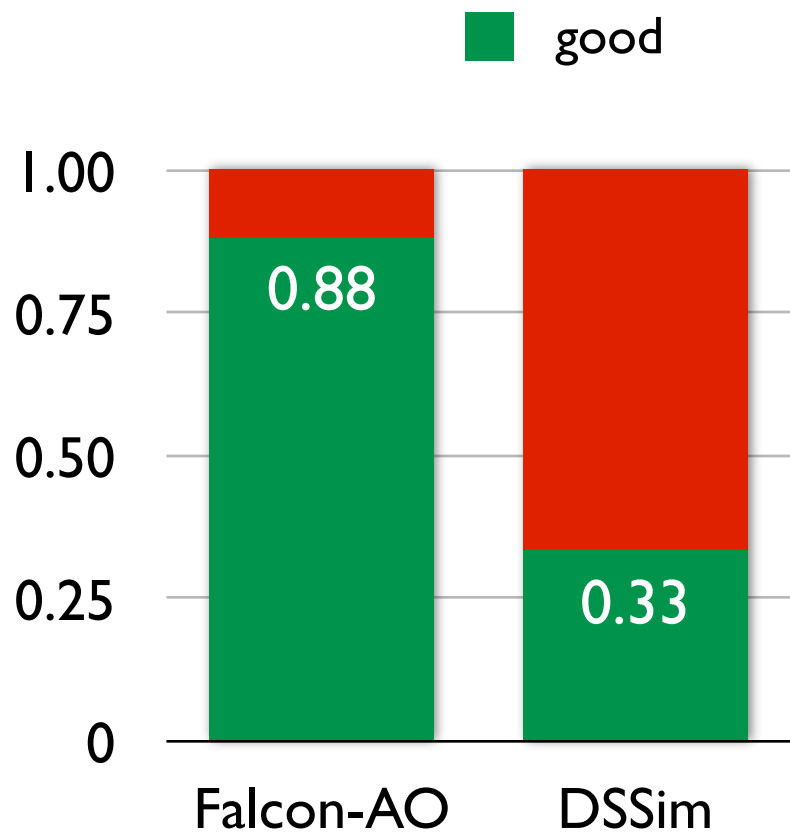




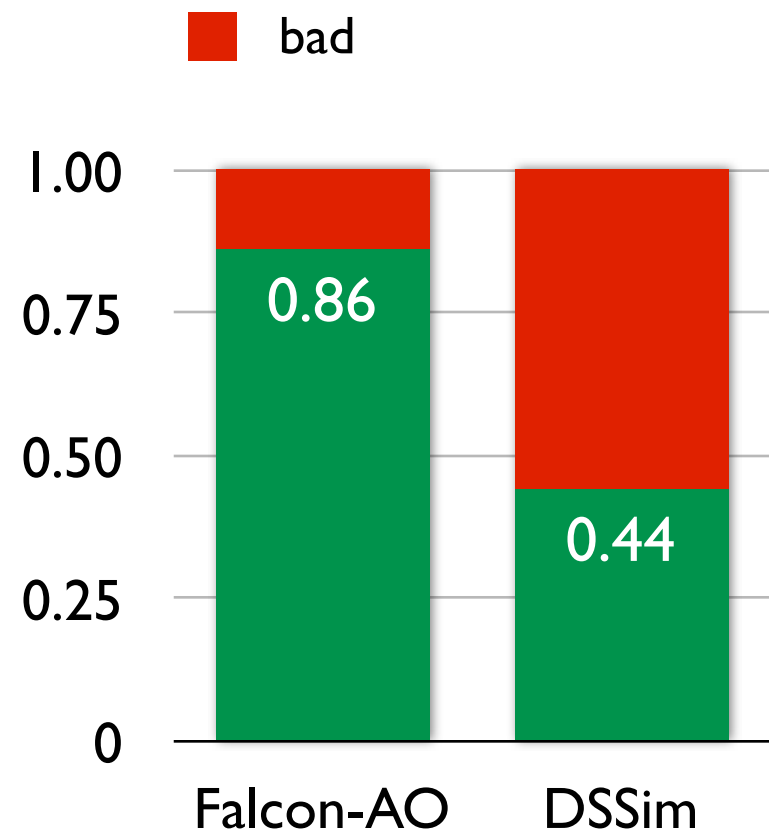
# precision of food safety mappings (15%, 15%)



# overall precision (100%, 100%)

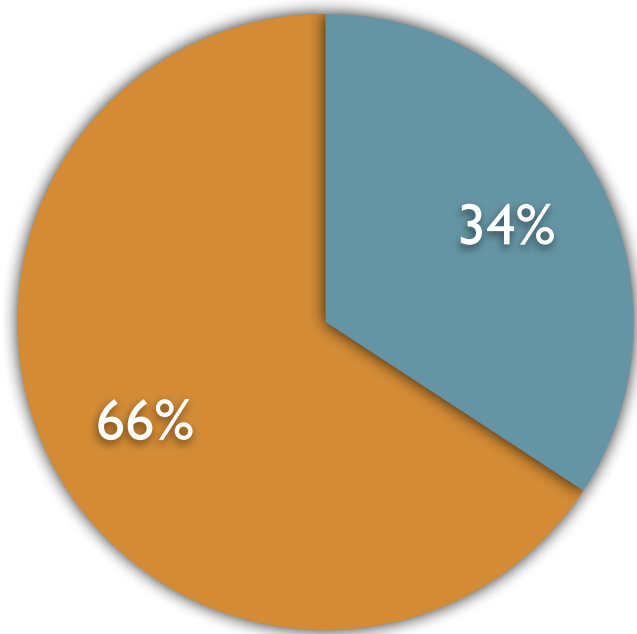


GEMET-AGROVOC

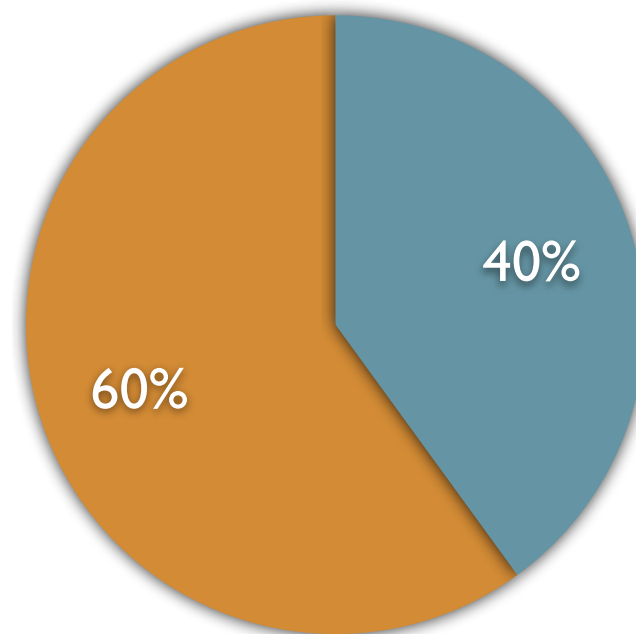


GEMET-NALT

# gold standard relations

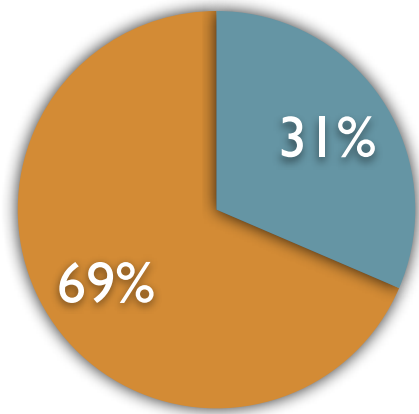


GEMET-AGROVOC

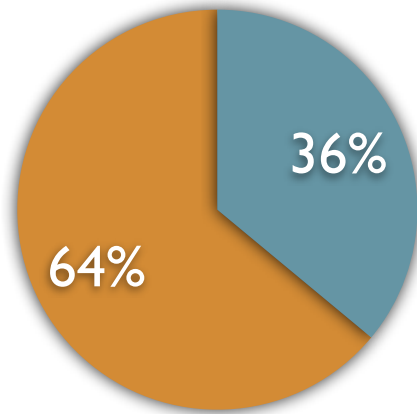


GEMET-NALT

# gold standard relations



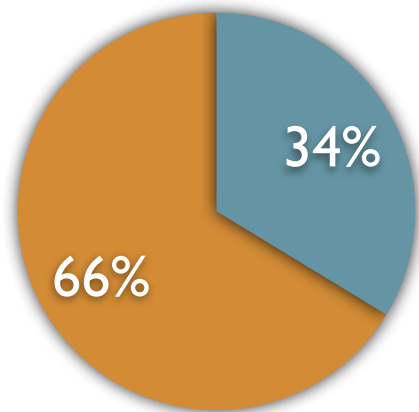
Agriculture



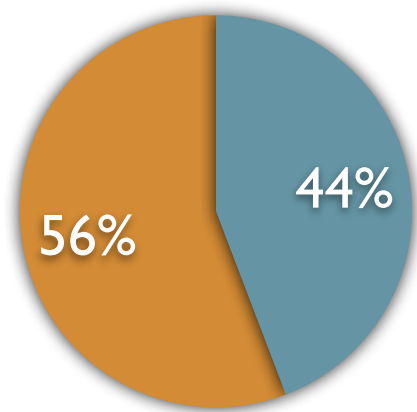
Geology

**GEMET-AGROVOC**

● broad & narrowMatch  
● exactMatch



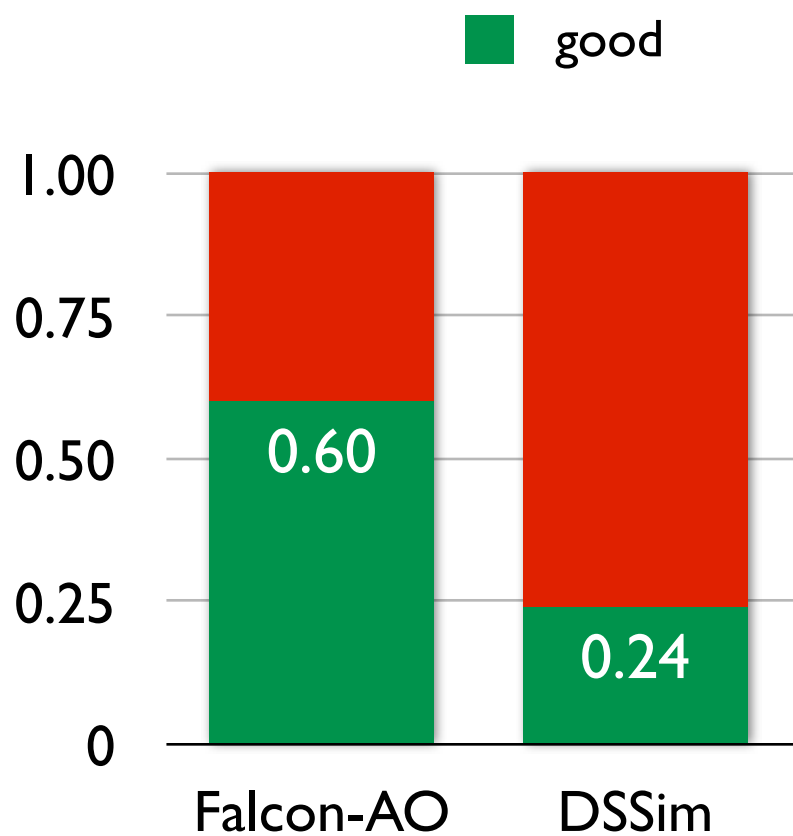
Agriculture



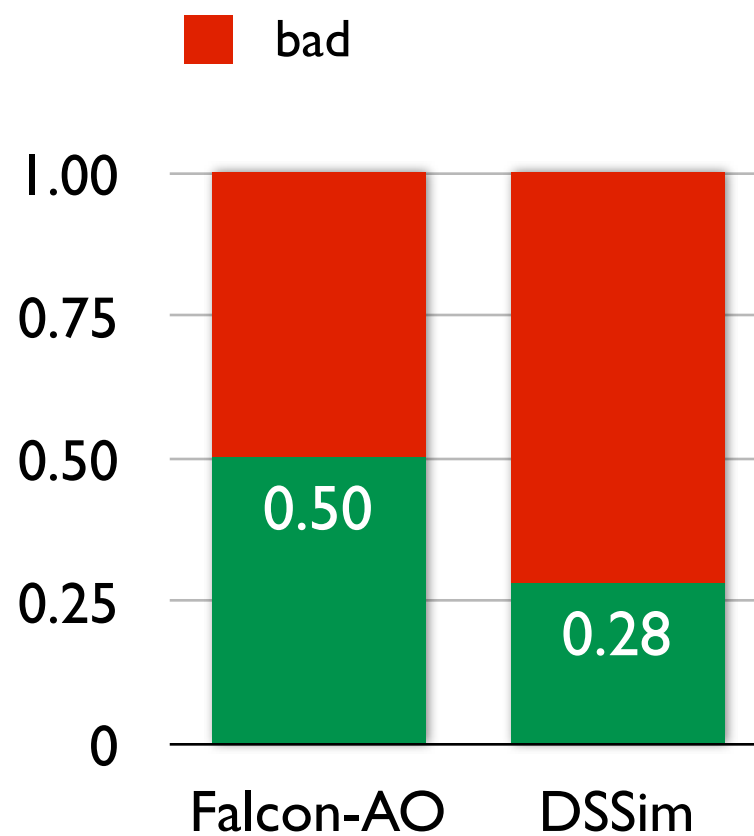
Geology

**GEMET-NALT**

# recall of only exactMatch



GEMET-AGROVOC



GEMET-NALT

# recall of exact, broad & narrowMatch

