

Relevance-based Evaluation the OAEI 2007 food task revisited

Willem Robert van Hage
Vrije Universiteit Amsterdam
wrvhage@few.vu.nl

Hap Kolb
TNO Science & Industry
hap.kolb@tno.nl

Guus Schreiber
Vrije Universiteit Amsterdam
schreiber@cs.vu.nl

1 Gather hot topics



2 Gather documents



3 Manually Match Metadata

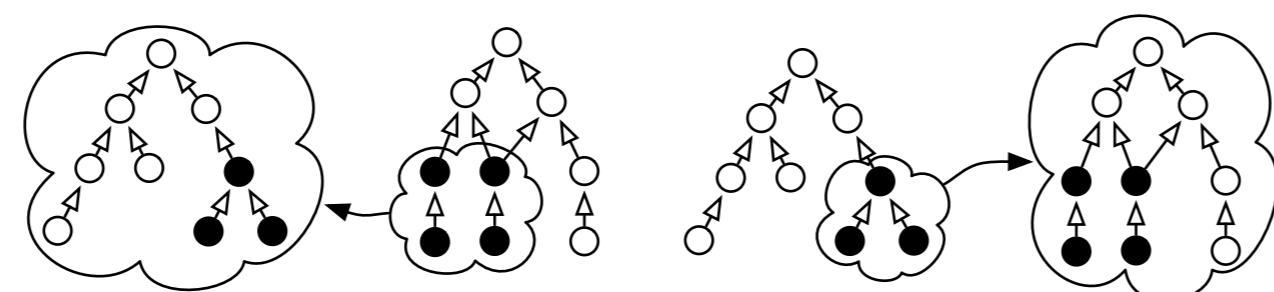
Record number US9307168
Titles Biofuels: January 1986-August 1992
Personal Authors Miller, E.K. (D.C. Reference Center, National Agricultural Library, Beltsville, MD)
Publication Date Sep 1992
AGRIS Subj. Cat. Renewable energy resources
AGROVOC Terms BIOMASS, FUELS, FUEL CROPS, ENERGY RESOURCES, BIOMASSE, COMBUSTIBLE, PLANTE ENERGETIQUE, RESSOURCE ENERGETIQUE, BIOMASA, COMBUSTIBLES, CULTIVOS ENERGETICOS, RECURSOS ENERGETICOS
Other subjects CROPS, ENERGY SOURCES, NATURAL RESOURCES, PLANTE DE CULTURE, RESSOURCE NATURELLE, SOURCE D'ENERGIE, CULTIVOS, FUENTE DE ENERGIA, RECURSOS NATURELES, BIBLIOGRAPHIES
Language En
Notes Bibliography
Type Bibliography, Non-Conventional
Collation 69 p.
Availability NAL/USDA (United States of America)
Availability number 9307168
Serial Title Quick bibliography series - U.S. Department of Agriculture, National Agricultural Library (U.S.) (USA)
Serial ID - ISSN 1052-5378
Serial number no. 92-63

FAO AGRIS/CARIS

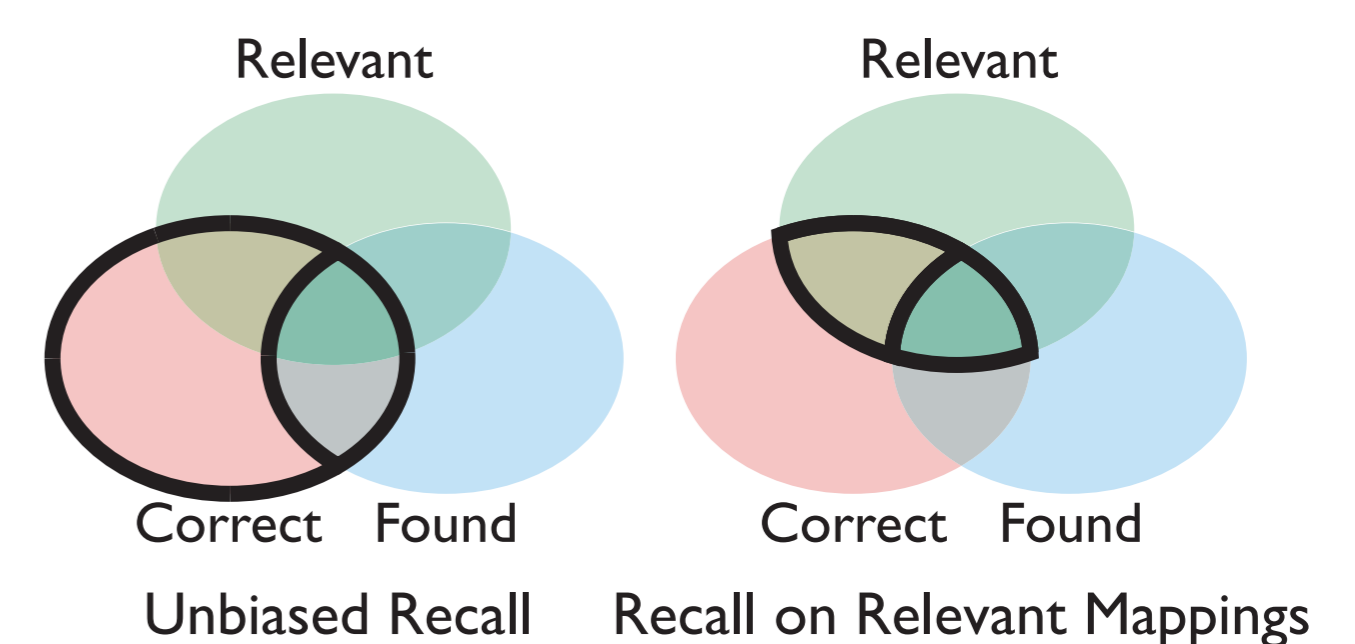
Biofuels--January 1986-August 1992.

Title: Biofuels--January 1986-August 1992.
Author: Miller, E.K.
Found In: Quick bibliography series - U.S. Department of Agriculture, National Agricultural Library (U.S.). Quick Bibliogr Ser U S Dep Agric Natl Agric Libr U S Sept 1992. (92-63) 69 p. ISSN 1052-5378
Note(s): Bibliography.
NAL Subject(s): biomass
fuels
energy resources
bibliographies
Other Subject(s): fuel crops
Subject Code(s): P120

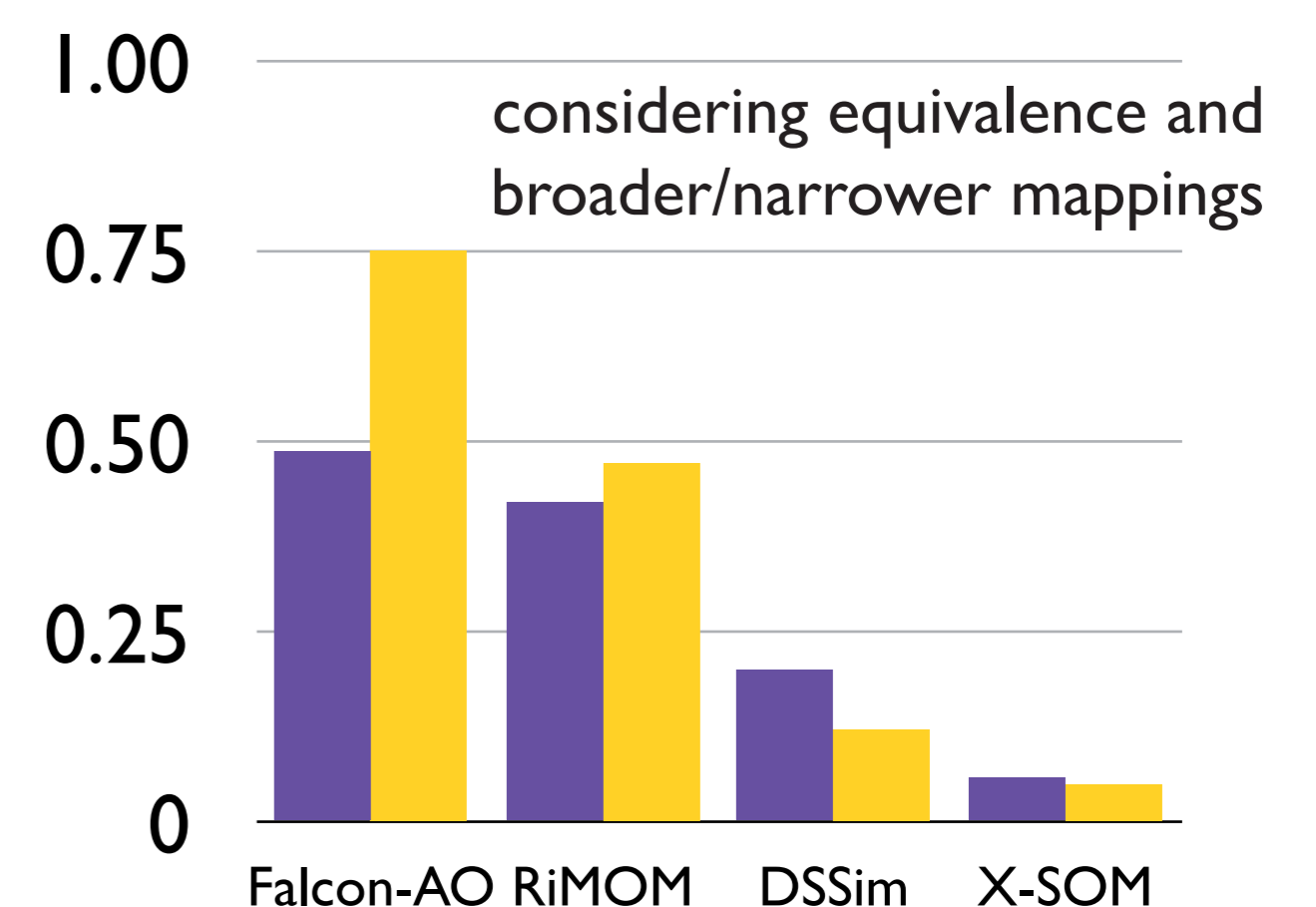
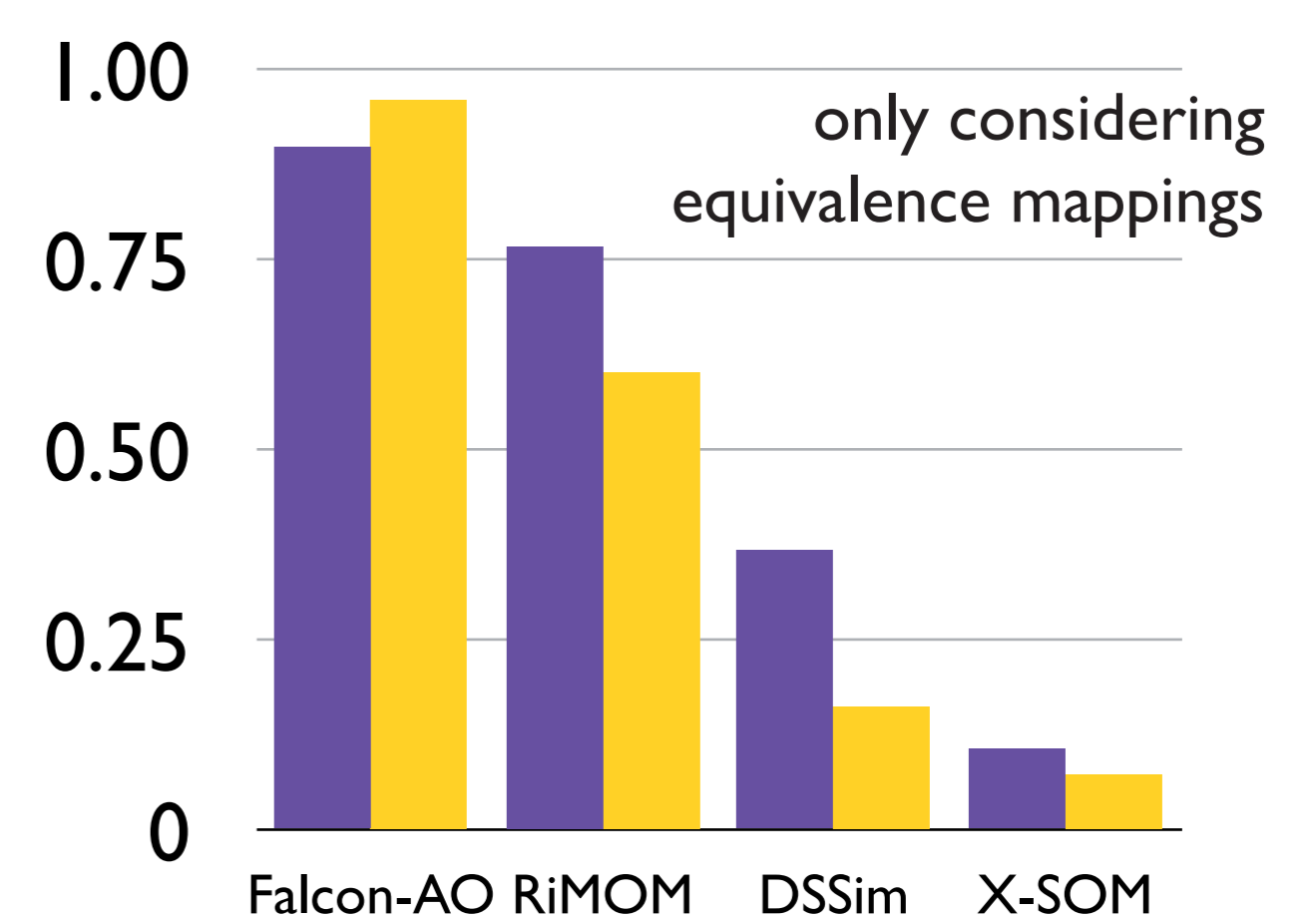
USDA AGRICOLA



4 Compute Recall on Relevant Sample



5 Results



Legend:
■ OAEI 2007 Recall
■ Relevance-based Recall

discussion

- The OAEI 2007 system ranking is consistent with the relevance-based second opinion.
- Especially “easy” mappings were needed to find the relevant mappings.
- The relevance-based results exaggerate the differences between the systems, because the best systems are good at finding all “easy” mappings first.
- It matters **which** correct mappings you find.
- A **typical** case is seldom an **average** case. Uniform sampling estimates **average case** behavior, not necessarily **typical** behavior.