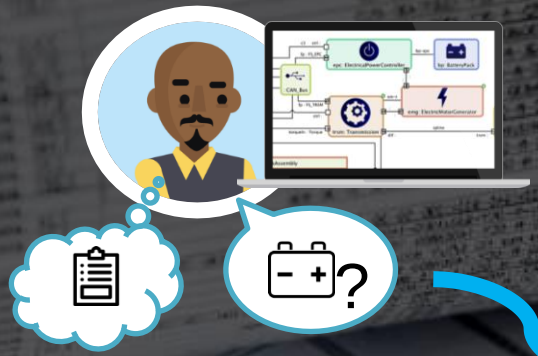
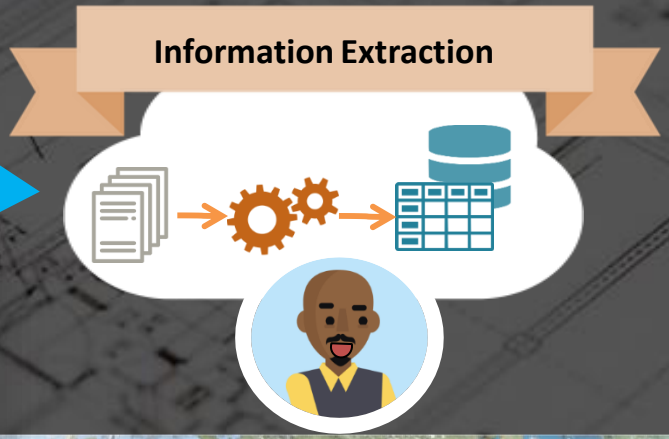
Towards Learning from User Feedback for Ontology-based Information Extraction

Kobkaew Opasjumruskit
German Aerospace Center
Institute of Data Science

kobkaew.opasjumruskit@dlr.de

Scenario





Information Extraction from Data Sheets

- Entity recognition → Either too generic or too specific

The ASTRO APS has been designed with compact dimensions, low mass, and low power consumption. Since July 2013, the ASTRO APS has been accumulating space heritage operating flawlessly on board of Alp

Jena-Optronik GmbH · Otto-Eppenstein-Straße 3 · 07745 Jena · Germany
Phone +49 3641 200-110 · **Fax** +49 3641 200-222
Email sales@jena-optronik.de · **Web** www.jena-

DBpedia Spotlight

The Jena-Optronik ASTRO APS is an Autonomous Star Sensor with the most advanced radiation hard CMOS Active Pixel Sensor detector technology for long-term missions on Telecom, Science and Earth Observation satellites.
space for success
The ASTRO APS has been designed with compact dimensions, low mass, and low power consumption.
Since http://dbpedia.org/resource/American_Society_for_Radiation_Oncology
space heritage operating flawlessly on board of Alphasat.

OPEN CALAIS

Picture Alphsat © ESA
Jena-Optronik Gm
Phone +49 3641 2
Email sales@jena-
Size & Mass
Dimensions 154 m
Mass approx. 2 kg

ESA | Contact Details | Company Technology

RELATION
CONTACT DETAILS

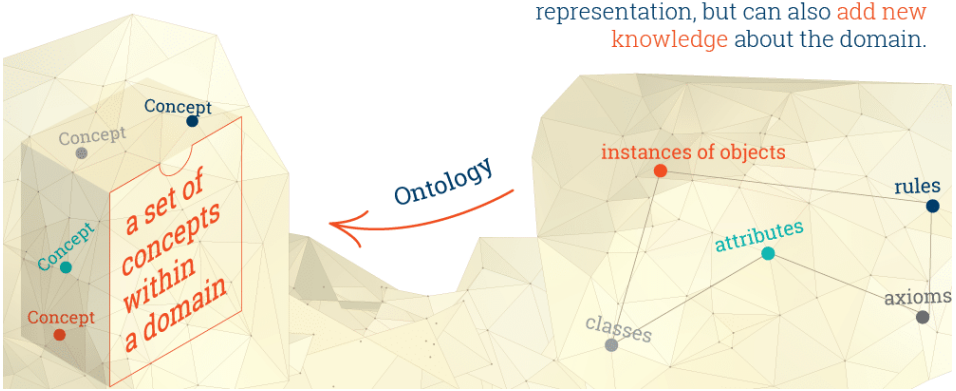
telephone	+49 3641 200-110
emailaddress	sales@jena-optronik.de
url	www.jena-optronik.com

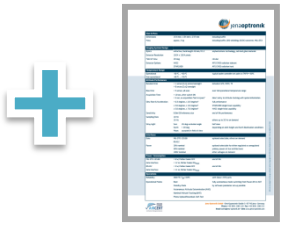
- Domain knowledge → Ontology

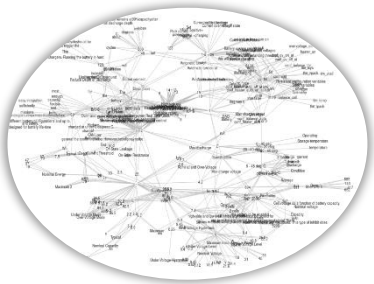
DLR

Semantic Knowledge - Ontology

Ontologies do not only introduce a sharable and reusable knowledge representation, but can also add new knowledge about the domain.

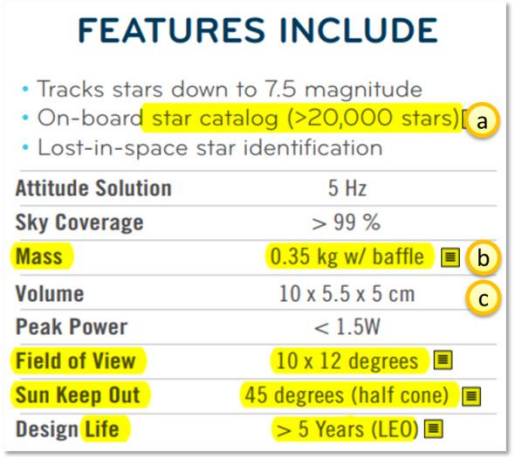






https://www.ontotext.com/knowledgehub/fundamentals/

Challenge - Key-value pairs detection



FEATURES INCLUDE

- Tracks stars down to 7.5 magnitude
- On-board star catalog (>20,000 stars) a
- Lost-in-space star identification

Attitude Solution	5 Hz
Sky Coverage	> 99 %
Mass	0.35 kg w/ baffle b
Volume	10 x 5.5 x 5 cm c
Peak Power	< 1.5W
Field of View	10 x 12 degrees
Sun Keep Out	45 degrees (half cone)
Design Life	> 5 Years (LEO)

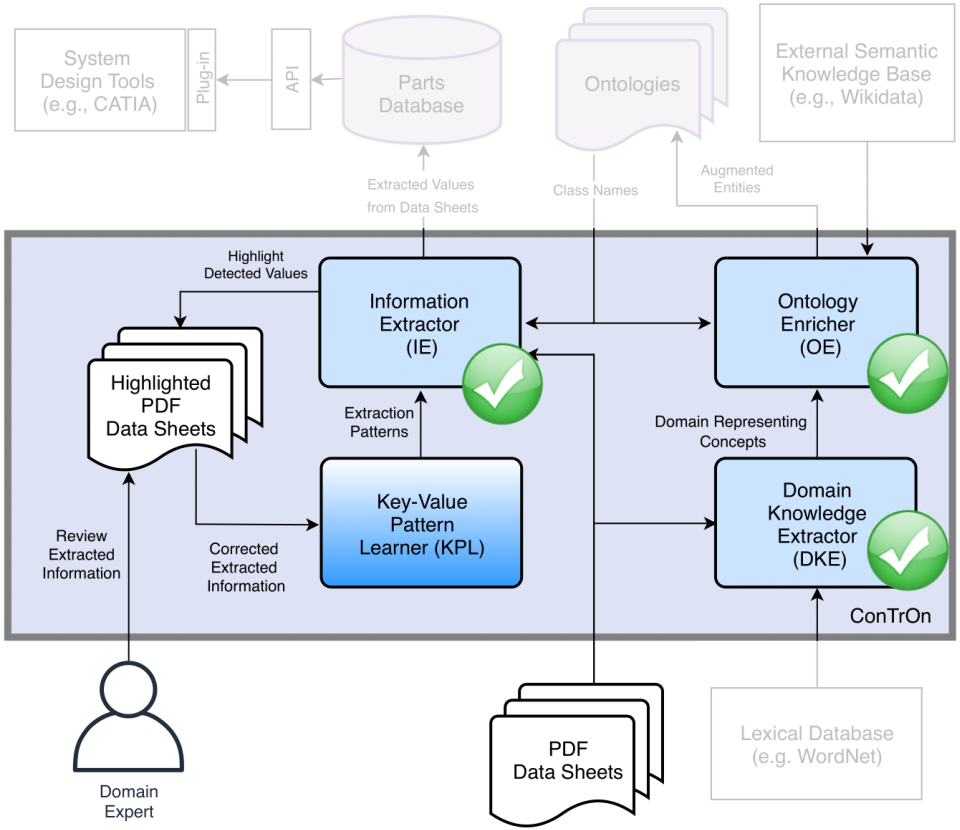
a) Incorrectly detected data

b) Correctly identified

- Unit and condition can be mixed up

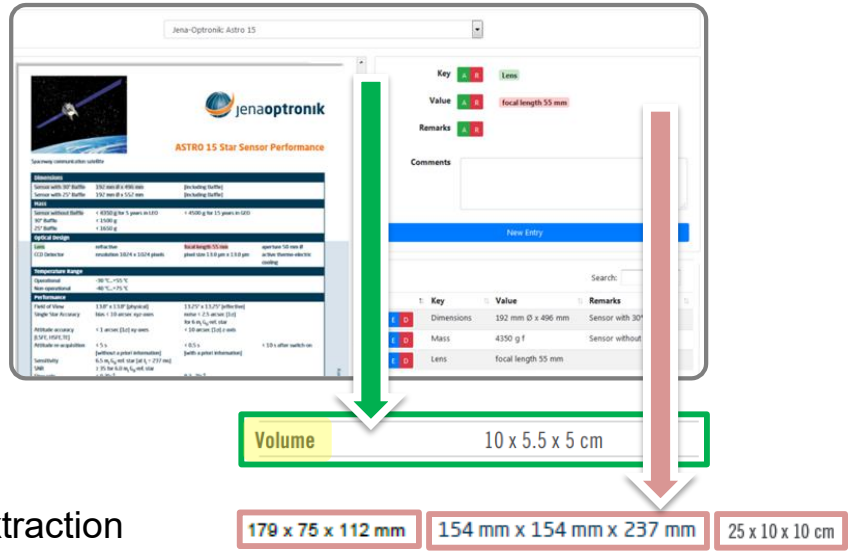
c) Undetected information

System Overview



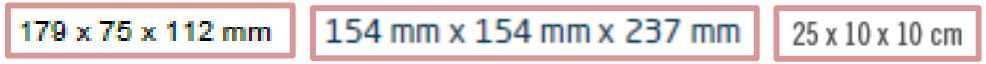
Key-Value Pattern Learner

- Users provide feedback via a tool



- Key → Add to the ontology
- Value → Improve the information extraction

Pattern Example



- Learned pattern will be used by IE (Information Extractor)
  - <number>+" x "+<number>+" x "+<number> + <unit>
  - <number> <*unit*>+" x “+<number> <*unit*>+" x “+<number> <unit>

- Verify the pattern
  - Apply the rules to the information extractor
  - Compare the extracted result again with the user’s feedback
  - Choose the pattern that yield the minimum error

Summary

- GOAL: Information extraction from data sheets

- Automatically enrich existing ontologies
  - ✓(I) Ontology Enrichment

- Discover more information from data sheets
  - ✓(II) Domain Knowledge Extraction
  - ✓(III) Ontology-based Information Extraction

- Challenge: Correctness & Completeness of Extracted Information
  - ❑(IV) Key-Value Pattern Learner (from user feedback)

Thank you for your attention!