

Semantic Web – Etapa 1

# Ontology Based Approach for Online Customers Reviews Analysis

Presented by: Gheorghe Alina, ISI

# Ideea Proiectului



**Integrarea între ontologie și sistemul de procesare a limbajului natural** în extragerea datelor privind recenziile clienților pe website-urile de e-commerce.

**Utilitatea** : aplicarea pe datele extrase de pe mai multe website-uri.

**Recenziile sunt considerate a fi informații textuale care conțin o evaluare a unui produs.**

O recenzie include ratingul acordat de către cumparator și încrederea sa în produs atunci când evaluează contribuția.

# Functionalitati

Principala functionalitate : obtinerea polaritatii opiniilor clientilor din ratingul si review-ul dat produselor de pe website-uri de tip e-commerce, precum si a aspectelor subliniate in aprecierea unui produs.

Modul Crawling	Modelare Date	Modul de Afisare	Modul NLP
<ul style="list-style-type: none"><li>• Adaptat fiecarui site: Decathlon, Emag, Altex, Notino</li><li>• Va extrage toate structurile necesare popularii grafului de cunostinte:<ul style="list-style-type: none"><li>◦ Website sursa</li><li>◦ Cod produs</li><li>◦ Titlu</li><li>◦ Categorie</li><li>◦ Rating</li><li>◦ Lista review-uri</li></ul></li><li>• Salvare csv</li></ul>	<ul style="list-style-type: none"><li>• Tripletele vor fi create cu RDFS</li><li>• Stocarea, parsarea, crearea claselor si proprietatilor cu frameworkul GraphDB</li><li>• Interogari: SPARQL</li></ul>	<ul style="list-style-type: none"><li>• Portal web minimal pentru<ul style="list-style-type: none"><li>◦ posibilitatea interogarii grafului de cunostinte</li><li>◦ Afisarea rezultatelor</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Model de limabj pentru extragerea aspectelor din review-uri si adaugarea lor in graful de cunostinte</li></ul>

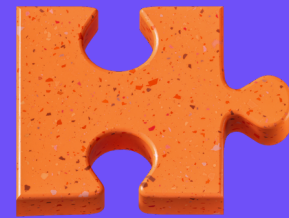




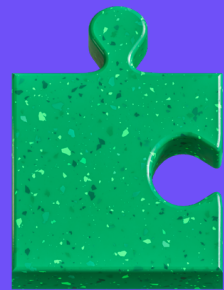
# Intrebari preconizate



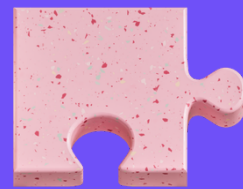
Ce categorie mare de produse are cel mai bun rating?



Ce sub-categorie de produse (cea mai specifica) a strans cele mai multe review-uri negative?

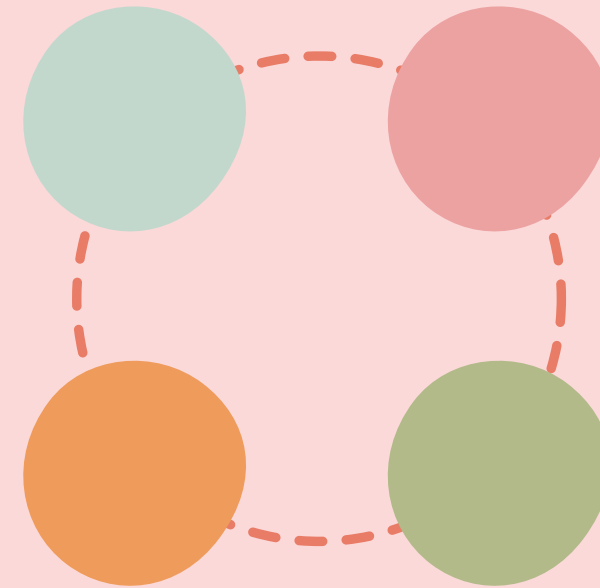


Care e ierarhia de categorii in care se incadreaza o instanta a unui produs?



Care este o distributie a polaritatilor opiniilor utilizatorilor despre un anumit produs?

# Folosirea Grafului de Cunoastinte



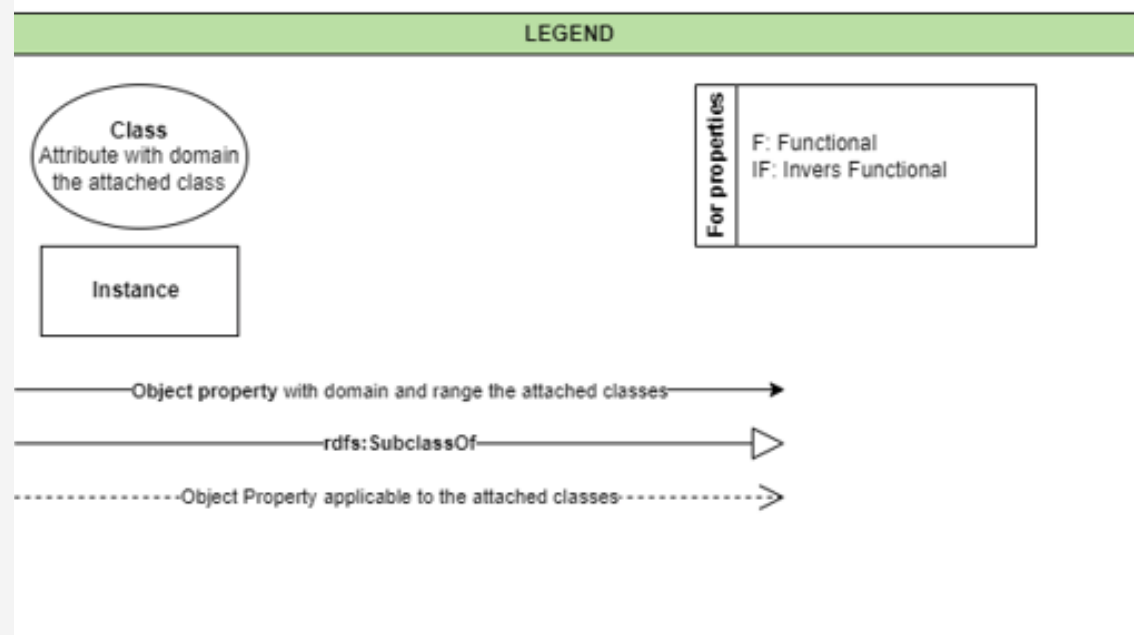
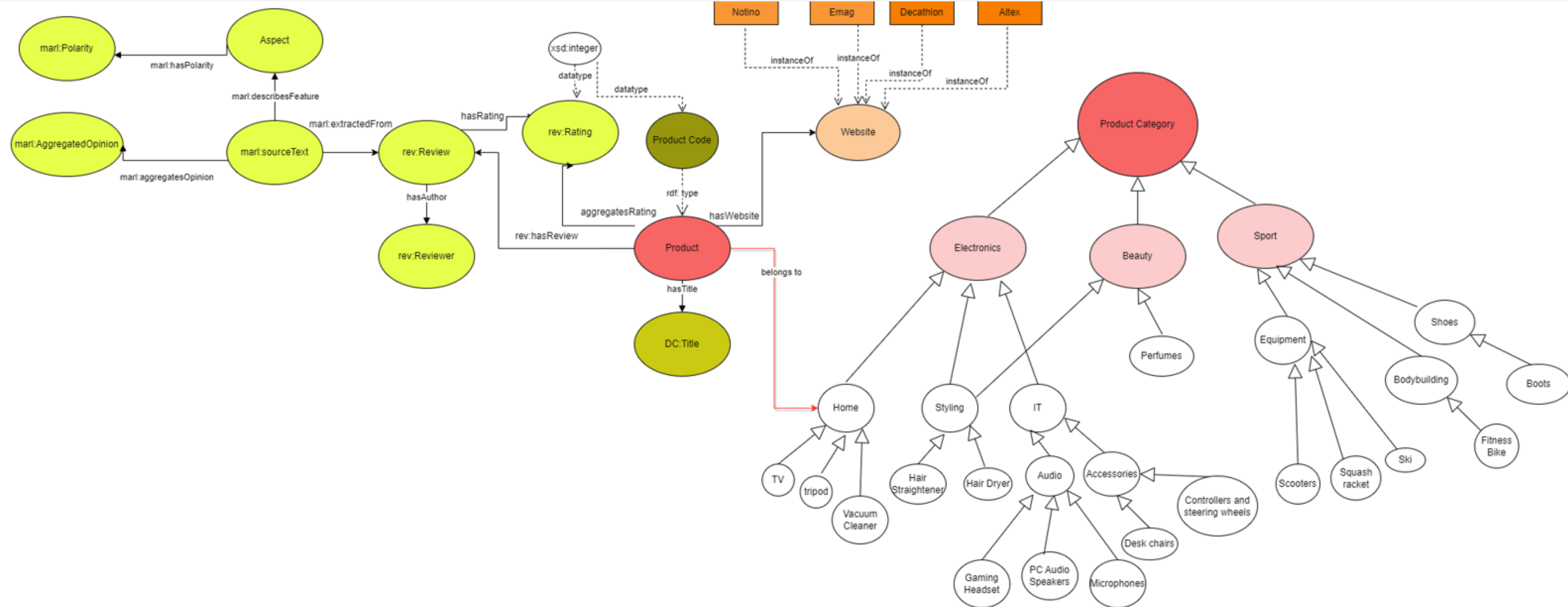
## KG vs Structuri de date relationale

- Conservarea intelesului logic al datelor
- Renuntarea la structurile relationale prin normalizare FN1/FN2/FN3
- Reutilizarea altor ontologii deja existente: REV & MARL
- Guvernarea principiului Open World Assumption

## Multiple surse de date

- Alegerea categoriilor de produse

Pentru a putea reduce bias-ul intalnit in recenzii sunt alese cateva categorii specifice din diferite surse populare



**VOCABULARY**

Vocabulary:

Imported vocabularies: <https://vocab.org/review/> (rev) / <http://purl.org/spar/fabio> , <https://www.qsi.upm.es/ontologies/marl> (marl)

Namespaces referenced: rev, rdfs, xsd



# Seturi de date primare



## Website-uri pentru crawling

- **Emag:** <https://www.emag.ro/>
- **Altex:** <https://www.altex.ro/>
- **Notino:** <https://www.notino.ro/>
- **Decathlon:** <https://www.decathlon.ro/>

## Alinierea cu ontologii existente:

- **MARL:** <https://www.gsi.upm.es/ontologies/marl/>  
([http://marl.gi2mo.org/?page\\_id=1](http://marl.gi2mo.org/?page_id=1))

Ontologie conceputa pentru a adnota și descrie opiniile subiective exprimate pe web sau în anumite sisteme de informații.

- **REV:** <https://vocab.org/review/> – review vocabulary

Vocabular conceput pentru exprimarea recenziilor și a evaluărilor utilizând cadrul de descriere a resurselor.

# Potentiale tehnologii

- ✓ Backend & Frontend : Python
- ✓ Ontologie: RDFS (definire triplete)
- ✓ Ontologie: GraphDB (creare, vizualizare, operare)
- ✓ Modelele de limbaj: Pytorch





# Bibliografie

- ✓ <https://www.sciencedirect.com/science/article/abs/pii/S0166361519301836>
- ✓ <https://sparontologies.github.io/fr/current/fr.html>
- ✓ <https://patentimages.storage.googleapis.com/7c/31/02/e4f58c122e4f12/US20160180437A1.pdf>
- ✓ [https://link.springer.com/chapter/10.1007/978-981-15-1097-7\\_8](https://link.springer.com/chapter/10.1007/978-981-15-1097-7_8)

