- **Semantic similarity:** This measurement provides a distance metric between a traditional ontology augmented with an influence network and the text data in fashion domain in terms of "meanings" they express. We project both labels (class names O_c and property names O_p) in the ontology and the tokens in the text corpus D in the same vector space (e.g., using word2vec). Then we compute the overall similarity based on all labels' distances weighted by their importance within the network:

$$\sum_{c_i \in O_c} \sum_{t \in c_i} Sim(t, D) * \theta(c_i) + \sum_{p_i \in O_p} \sum_{t \in p_i} Sim(t, D) \quad (4)$$

The importance score $\theta(c_i)$ is a normalized score $[0, 1]$ and can be defined depending on the context and usage. For example, $\theta(c_i)$ refers to how knowledgeable a network is w.r.t. class c_i, which can be approximated by the cumulative distribution function of its number of attributes u_i

$$\theta(c_i) = \sum_{u_m < u_i} P(U = u_m) \quad (5)$$

4.1.2 Task-specific expressiveness

This is a task-driven measurements to quantify how expressible the ontology is compared to user's mental representation in the context of a task. Here we use information retrieval in the fashion domain as an example task, and developed statistical measures of "expressiveness".

- **Query concept recall:** Given a query stream like "natural fabric button down from banana republic", we derive a mapping between concepts that appeared in the query and concepts encoded in the ontology. Specifically, we ask expert judges to identify important classes $\{c_i\}$ in the query. For each class being detected, we ask them to map it to the most similar label in the ontology. Then we compute the recall of concepts in the query as $\frac{\#\text{ concepts mapped to ontology}}{\#\text{ concepts detected}}$

- **Search result ranking:** We use rankings of search results for a given query as a proxy of "golden standards". Then the Normalized Discounted Cumulative Gain (NDCG) metric can be adapted to measure the ontology's relevancy to search quality. Specifically, the "gain" is quantified by the ontology's recall of concepts in search results. We examine the top K returned search results. For each of them D_p at position p, we compute the ontology's document concept recall, which is then discounted by its logarithmic rank $log(p)$.

$$\sum_{p=1}^{K} \frac{Recall(O, D_p)}{log(p)} \quad (6)$$

4.2 Applications

4.2.1 Web Data Markup through schema.org

There are massive data sources on the Web (including mobile applications). However, they are mostly unstructured, and there is no common vocabulary which facilitates collective curation of domain knowledge. Schema.org markup has been the major adoption for web data (about 31.3% of all pages by Dec 2015[5]) and is used by a variety of high traffic applications like search engines and news portal. Although it contains vertical specific schemas such as movies, music, medical and products, schema that can represent fashion content is absent.

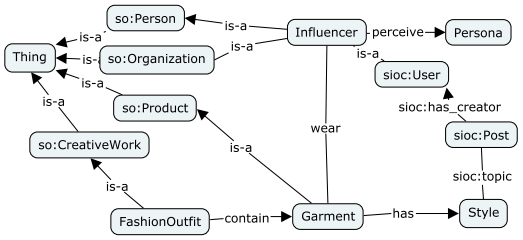

Figure 4: Integration with schema.org

In Figure 4 we illustrate that the proposed ontology framework can easily adapt to a lightweight ontology and integrated as an external extension of the core schema.org vocabulary, while also linking to other relevant common vocabularies such as the *GoodRelations* for E-commerce[6], *SIOC* for influence mechanisms on the social Web [3].

4.2.2 Consumed by machine learning systems

The data represented in a subjective influence network as proposed here could be used for a variety of different data analysis and processing efforts, including the following types:

- **General machine learning problems:** The knowledge base represented by a populated influence network would contain instances associated with high-quality categorical types, which provide labeled data to train models for entity recognition. Also, the edges on their own in the network contain both numeric and categorical features which can be used in whole-network modeling experiments.

- **Fashion data retrieval:** The integration with `schema.org` enables community content publishers to explicitly annotate their posts with their perceived subjectivity of fashion contents, which are basic building blocks of a crowd-sourcing system. As a result, the marked up Web data in return allows for information organizers such as search engines to index rich contents and answer queries which contain both entities and subjective projections.

- **Recommender systems:** The taxonomy defined in the ontology provides a perfect complement to recommendations learnt in a bottom-up fashion. Therefore, it could be a very useful approach to deal with data sparsity situations such as cold start problems.

5. CONCLUSIONS

In this paper we propose a new ontological augmentation in the fashion domain, which represents subjective feature information as an influence network. Because fashion (just like art, music or languages) strongly contains subjective information (cultural phenomena which are not designed nor