## **Output:**

 $brands = top_{10}(\mathcal{S}) \mid items = top_{10}(\mathcal{S})$ 

 $styles = top_{10}(\mathcal{S}) \mid patterns = top_{10}(\mathcal{S}) \mid materials = top_{10}(\mathcal{S})$ 

## Ambiguity Resolution by ProBase (Knowledgebase) API lookup:

 $\forall o_i \in \mathcal{O} \ \mathcal{S}(o_i) := probase(o_i) \cdot \mathcal{S}(o_i)$ 

Semantic Similarity (Cosine similarity) with Ontology Items  $\mathcal{O}$  (aka Clustering in  $\mathbb{R}^{300}$ ):

 $\forall o_i \in \mathcal{O} \quad \vec{o_i} \in \mathcal{W}$  $\sum_{t \in \mathcal{T} \vec{t} \in \mathcal{W}} \mathcal{S}(o_i) = \mathcal{S}(o_i) + \alpha \cdot \cos(\vec{o_i} \vec{t})$ 

Syntactic Similarity (Levenshtein distance) with Ontology Items  $\mathcal{O}$ :  $\forall o_i \in \mathcal{O}$ 

 $\sum_{t \in \mathcal{T}} \mathcal{S}(o_i) = \mathcal{S}(o_i) + \beta \cdot lev(o_i t)$ 

Linguistic Normalization:  $\mathcal{T} = \text{lemmatize} \circ \text{stopword-removal} \circ \text{social-media-tokenizer}(\mathcal{D})$ 

Input:

**Syntactic Normalization:**  $\mathcal{D} = \text{lower-case} \circ \text{remove-urls}(\mathcal{D})$ 

Bag of words document:  $\mathcal{D}$ Word embeddings in  $\mathbb{R}^{300}$ :  $\mathcal{W} \in \mathbb{R}^{300 \times |V|}$ Scored ontology categories  $S := \{0..\}$ 

Tokens  $\mathcal{T} := \emptyset$