DiG is a walk-up-and-use system that lets people make informed purchasing decisions about products.

DiG recommends products to users based on how they plan to use the product.

The interface uses information extracted automatically from sentences in online customer reviews. Statistical methods that cluster and find patterns are used to get detailed information about product features and uses.

**Problem: fine-grained extraction**
- Need to extract a list of a product’s features and uses from customer reviews with high precision and low redundancy
- Cannot use existing web-scale algorithms (not enough data)
- Need to accurately evaluate customers’ overall opinion of a product feature from customer reviews

**Solution: Statistical and Linguistic Information**
- Use human knowledge to:
  - Map product uses to features
  - Provide a few examples of product features
  - Annotate adjectives with sentiment
- Use grammatical structure to:
  - Filter out incoherent extractions
  - Characterize extracted items for grouping
  - Accurately identify opinion-giving adjectives
- Use statistical methods to:
  - Filter out infrequent and imprecise extractions
  - Cluster extracted items into groups
  - Predict the opinion polarities of adjectives

**Data: Digital Cameras**
- Standard Product Specifications
- Product Attributes
- Reviews
  - 168,638 reviews for 3700 cameras from Amazon
  - Over 1.1 million sentences

**Processing Reviews**

### Product Feature Extraction

1. **Learn extraction patterns**
   - digital camera features a < > and ...
   - the < > on this camera ...
   - the camera’s < > is ...
   - camera has < >, which ...
2. **Pattern-match to get candidates**
   - digital camera features a “macro setting” and ...
   - the “image stabilization” on this camera ...
   - the camera’s “best selling point” is ...
   - camera has “grainy images at night”, which ...
3. **Identify reliable product features**
   - SVM classifier
   - Features based on web-PMI
4. **Cluster product features**
   - Cluster noun sequences based on noun-adj distributions
   - Filter with reliable product features

### Opinion Mining

5. **Classify adjectives: PMI features**
   
   - Web PMI = log hits(a, b)/hits(a)hits(b)
   - +1: “good”, “excellent” (+1.98, 0.27), “bad” 0.21
   - -1: “boring”, “poor” (-1.98, 0.27), “sharpen” (-0.21)

6. **Evaluate opinion: use adjectives**
   - image quality: grainy (+1), crisp (+1), clear (+1) ...
   - battery life: long (+1), good (+1), excellent (+1) ...
   - We use beta-binomial model to smooth estimates when there are only a few adjectives for a product feature

7. **Extract summarizing sentences and terms**

**Product Use Extraction**

8. **Pattern-match to get candidates**
   - <picture term> <prep phrase> <noun phrase>

9. **Filter candidates**

**Integrating Analysis Results into an Interactive UI**

- Allow users to describe how they will use a product
- Bridge between chosen uses and resultant features
- Compare a chosen set of cameras

Users can specify types of pictures, activities, and how they use their photos