

Chapter 9. E-Commerce and Retail

*Today's new marketplaces must nurture
and manage perfect competition to thrive.*
—Jeff Jordan, Andreessen Horowitz

In today's world, e-commerce has become synonymous with shopping. An enriched customer experience compared to what a physical retail store offers has fueled this growth of e-commerce. Worldwide retail e-commerce sales in 2019 were \$3.5 trillion and are projected to reach \$6.5 trillion by 2022 [1]. Recent advancements in ML and NLP have played a major role in this rapid growth.

Visit the home page of any e-retailer, and you'll find a lot of information in the form of text and images. A significant portion of this information consists of text in the form of product descriptions, reviews, etc. Retailers strive to utilize this information intelligently to deliver customer delight and build competitive advantage. An e-commerce portal faces a range of text-related problems that can be solved by NLP techniques. We saw different kinds of NLP problems and solutions in the previous section (Chapters 4 through 7). In this chapter, we'll give an overview of how NLP problems in the e-commerce domain can be addressed using what we've learned in this book so far. We'll discuss some of the key NLP tasks in this domain,

including search, building a product catalog, collecting reviews, and providing recommendations.

Figure 9-1 shows some of these e-commerce tasks. Let's start with an overview of them.

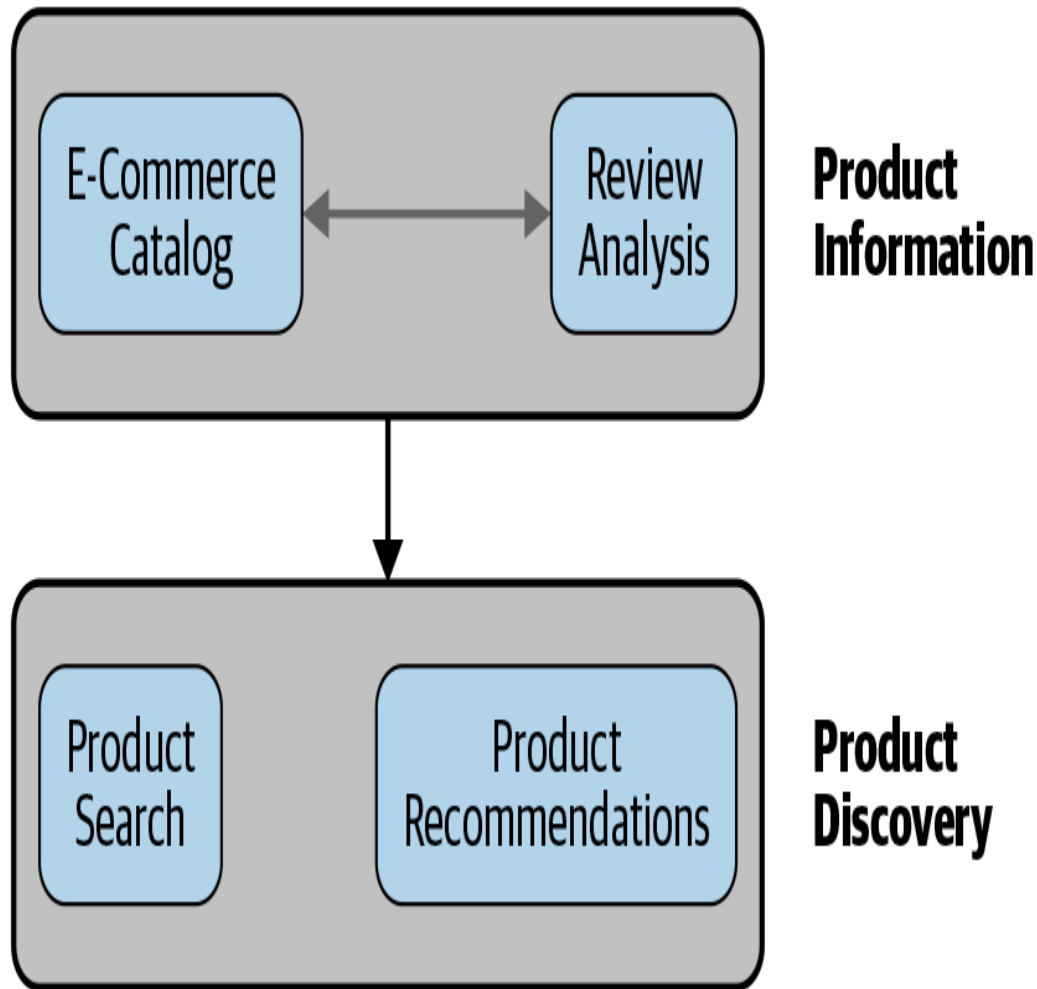


Figure 9-1. NLP applications in e-commerce

E-Commerce Catalog

Any large e-commerce enterprise needs an easy-to-access product catalog. A product catalog is a database of the products that the

enterprise deals or a user can purchase. This contains product description attributes as well as images for each product. Better product descriptions with relevant information help the customer choose the right product through the catalog. Such information can also help in product search and recommendations. Imagine a recommendation engine that automatically knows that you like the color blue! That's certainly not possible unless and until the engine notices that most of your recent purchases or searches were on apparel of the color blue. The first thing needed to achieve this is identifying that "blue" is associated with the products as a color attribute. Extracting such information automatically is called *attribute extraction*. Attribute extraction from product descriptions can guarantee that all the relevant product information is properly indexed and displayed for each product, improving product discoverability.

Review Analysis

The most notable part of an e-commerce platform is the user reviews section for every product. Reviews provide a different perspective of the product that cannot be obtained from the product attributes alone, such as quality, usability, comparisons with other products, and delivery feedback. All reviews may not be useful, or they might not come from trusted users. Further, it's hard to process multiple reviews for a given product manually. NLP techniques provide an overall perspective for all reviews by performing tasks such as sentiment analysis, review summarization, identifying review helpfulness, and so on. We saw one example of NLP for review analysis in Chapter 5

when we discussed keyphrase extraction. We'll see other use cases later in this chapter.

Product Search

Search systems in e-commerce are different compared to general search engines like Google, Bing, and Yahoo. An e-commerce search engine is closely tied to the products available and the different kinds of information associated with them. For instance, in a regular search engine, we're dealing largely with free-form text data (like news articles or blogs) as opposed to structured sales and review data for e-commerce. We might search for "red checkered shirt for a wedding," and the e-commerce search engine should be able to fetch it. Similar forms of focused search can also be seen on travel websites for flight and hotel bookings, such as Airbnb and TripAdvisor. The specific nature of the information associated with each type of e-commerce business calls for a customized pipeline of information processing, extraction, and search.

Product Recommendations

Without a recommendation engine, any e-commerce platform would be incomplete. A customer likes when the platform intelligently understands their choices and suggests products to buy next. It actually helps the customer organize their thoughts about shopping and helps to achieve better utility. Recommendations of discounted items, same-brand products, or products with favorite attributes can really engage the customer on the website and make them spend more time. This directly increases the possibility of the customers buying those products. In addition to transaction-based recommendation

facilities, there is a rich set of algorithms that are developed based on product content information and reviews that are textual in nature. NLP is used to build such recommendation systems.

With this overview, we're all set to explore the role of NLP in e-commerce in more detail. Let's start with how it's used in building search for e-commerce.

Search in E-Commerce

Customers visit an e-commerce website to find and purchase their desired products quickly. Ideally, a search feature should enable the customer to reach the right product with the least number of clicks. The search needs to be fast and precise and fetch results that closely match customers' needs. A good search mechanism positively impacts the conversion rate, which directly impacts the revenue of the e-retailer. Globally, on average, only 4.3% of user search attempts convert to a purchase. By some estimates, 34% of results in search on the top 50 portals do not produce useful results [2], and there's often a large scope for improvement.

In Chapter 7, we discussed how general search engines work and where NLP is useful. However, for e-commerce, the search engine needs to be more fine-tuned to the business needs. Search in e-commerce is closed domain—i.e., the search engine typically fetches items from within the product information, rather than from a generic set of documents or content on the open web (like Google or Bing). The underlying product information is built on the product catalog, attributes, and reviews. Search works on different facets of this

information, like color, style, or category. This kind of search in e-commerce is generally called “faceted search,” which is the focus of this section.

Faceted search is a specialized variant of search that allows the customer to navigate in a streamlined way with filters. For example, if we’re planning to buy a TV, then we might look for filters like brand, price, TV size, etc. In e-commerce websites, users are presented with a set of search filters depending on the product. Figures [9-2](#) and [9-3](#) illustrate search in e-commerce through Amazon and Walmart.

The left-most section of both images depicts a set of filters (alternatively, “facets”) that allows the customer to guide their search in a way that matches their buying needs. In [Figure 9-2](#), we see a search for television models, so the filters show aspects such as resolution and display size. Along with such custom filters, there are also some general features that are valid for many such product searches, such as brand, price range, and mode of shipping, as shown in [Figure 9-3](#). These filters are explicit dimensions to perceive the product. And this guided search enables the user to arrange the search results on their own to get more control over shopping, rather than having to sift through a lot of results to get what they’re looking for.

Amazon Prime



Eligible for Free Shipping

☐ Free Shipping by Amazon

TV Display Size

☐ 32 Inches & Under

☐ 33 to 43 Inches

☐ 44 to 49 Inches

☐ 50 to 59 Inches

☐ 60 to 69 Inches

☐ 70 Inches & Up

Television Resolution

☐ 4K

☐ 1080p

☐ 720p

Television Feature

☐ Smart TV

☐ 3D

Electronics Device Model Year

☐ 2017

☐ 2016

☐ 2015

☐ 2014

Television Refresh Rate

☐ 60 Hz

☐ 120 Hz

☐ 240 Hz

Brand

☐ Samsung

☐ TCL

☐ Sceptre

☐ LG

☐ Sony

☐ Element

☐ VIZIO

☐ Avera

☐ Element Electronics



Sony XBR65X930E 65-Inch 4K HDR Ultra HD TV (2017 Model)

by Sony

\$798⁰⁰ ~~\$1,199.99~~ **prime**

FREE Shipping on eligible orders

★★★★☆ • 125

Amazon Certified: Works with Alexa

- Display Size: 55.0 inches
- Resolution: 4K
- Connectivity Technology: Built-in Wi-Fi
- Model Year: 2017
- Display Technology: LED



Sponsored ⓘ

Sony XBR65X930E 65-Inch 4K HDR Ultra HD TV (2017 Model)

by Sony

\$2,498⁰⁰ **prime**

FREE Shipping on eligible orders

★★★★☆ • 68

Amazon Certified: Works with Alexa

- Display Size: 65.0 inches
- Resolution: 4K
- Connectivity Technology: Built-in Wi-Fi
- Model Year: 2017
- Display Technology: LED

Best Seller



TCL 55S405 55-Inch 4K Ultra HD Roku Smart LED TV (2017 Model)

by TCL

\$399⁹⁹ ~~\$599.99~~ **prime**

FREE Shipping on eligible orders

More Buying Choices

\$399.99 (10 used & new offers)

★★★★☆ • 1,449

- Display Size: 55.0 inches
- Resolution: 4K
- Connectivity Technology: Built-in Wi-Fi
- Model Year: 2017
- Display Technology: LED



TCL 40S305 40-Inch 1080p Roku Smart LED TV (2017 Model)

by TCL

\$269⁹⁹ ~~\$289.99~~ **prime**

FREE Shipping on eligible orders

More Buying Choices

\$269.50 (18 used & new offers)

★★★★☆ • 2,264

- Display Size: 40.0 inches
- Resolution: 1080p
- Connectivity Technology: Built-in Wi-Fi
- Model Year: 2017
- Display Technology: LED

Figure 9-2. Faceted search on Amazon.com

Shipping & Pickup

☒ Show all

☐ 2-DAY SHIPPING

☐ SHIP TO HOME

☐ FREE PICKUP + DISCOUNT

☐ FREE PICKUP

☐ FREE PICKUP TODAY

Brand

☐ Apple

☐ Total Wireless

☐ Simple Mobile

☐ Straight Talk

[See more brands](#)

Price

\$ to \$ [Go](#)

☐ \$50 - \$100

☐ \$100 - \$150

☐ \$150 - \$200

☐ \$200 - \$250

Refine

Price

Top brands

Store availability

Sort

Best sellers



Best Seller



List \$199.00 **Save \$40.00**
\$159.00

Straight Talk Apple iPhone SE
32GB Prepaid Smartphone,
Gray

★★★★★ (58)

2-DAY SHIPPING
FREE PICKUP



\$799.00

Straight Talk Apple iPhone 8
Plus with 64GB Prepaid
Smartphone, Si...

★★★★★ (4)

2-DAY SHIPPING
FREE PICKUP



\$699.00

Straight Talk Apple iPhone 8
with 64GB Prepaid
Smartphone, Silver

★★★★★ (2)

2-DAY SHIPPING
FREE PICKUP

Best Seller



Was \$163.00 **Save \$73.01**
\$89.99

iPhone SC 8GB Sprint
(Locked)

★★★★★ (9)

2-DAY SHIPPING
FREE PICKUP

Rollback



Was \$199.00 **Save \$39.01**
\$159.99

Simple Mobile Apple iPhone
SE 32GB Prepaid
Smartphone, Gray

★★★★★ (4)

2-DAY SHIPPING
FREE PICKUP



\$799.00



List \$199.00 **Save \$39.01**
\$159.99



\$799.00



\$624.00



\$630.40

Figure 9-3. Faceted search on Walmart.com

These filters are the key that defines the faceted search. However, they may not always be readily available for all products. Some reasons for that are:

- The seller didn't upload all the required information while listing the product on the e-commerce website. This is typically the case when a new e-commerce business ramps up and aggressively promotes quick onboarding of various sellers. To achieve this, they often allow the sellers to list without having quality checks in place for the product metadata.
- Some of the filters are difficult to obtain, or the seller may not have the complete information to provide—for example, the caloric value of a food product, which is typically derived from the nutrient information provided on the product case. E-retailers don't expect this information to be provided by the seller, but it's crucial because it may capture important customer signals that are directly related to the conversation of that product sale.

Faceted search can be built with most popular search engine backends like Solr and Elasticsearch. Besides regular text search, different facet attributes are also added to the search query. Elasticsearch's DSL also comes with a built-in faceted search interface [3].

TIP

In an e-commerce setting, we also need to account for business needs other than relevance in terms of facets and text. For instance, products that are part of a promotion or sale may be bumped up in results. This can be built by utilizing features like Elasticsearch boosting.

Apart from search algorithms, there are many nuances associated with faceted search, and we'll focus on these for the rest of this chapter. The issues mentioned above relate to the problem we'll discuss in the next section: building an e-commerce catalog.

Building an E-Commerce Catalog

As we saw earlier in this chapter, building an informative catalog is one of the primary problems in e-commerce. It can be split into several subproblems:

- Attribute extraction
- Product categorization and taxonomy creation
- Product enrichment
- Product deduplication and matching

Let's take a look at each of these in this section.

Attribute Extraction

Attributes are properties that define a product. For example, in [Figure 9-2](#), we saw brand, resolution, TV size, etc., as relevant