Solr/Lucene for eCommerce

Solr is a leading open source enterprise search platform that is highly extensible with companies that have the technical knowhow to configure and build on it. Out of the box, it features full-text search, faceted search, and can search among abstract documents. It is highly reliable, scalable, and fault tolerant. Solr, however, is not built out of the box to accommodate eCommerce applications. An eCommerce search system must satisfy the customer by providing the most relevant purchasable items in the fewest possible clicks. Whether shopping for carpets, dresses, or soup, the customer makes a first level of selection such as a brand, pattern, and style and then further narrows down to a purchasable item by selecting color, size, or flavor. If all the possible purchasable variants are displayed the shopping experience is cluttered and confusing.

Grid Dynamics has the experience to help retailers build out Solr so it is optimized for eCommerce. We can take out of the box Solr or the SAP hybris Solr engine and extend those capabilities. There are four areas where Solr needs modification to provide retailers the tools for the best customer experience: Catalog, Customer, Merchandiser, and Stability and Elasticity.

catalog

An eCommerce catalog is deeply relational and usually represents hierarchies of collection->product->sku->price. Some searchable characteristics may belong to a product like brand, other characteristics may belong to a sku like size or color. To display accurate facets, including the numbers of products that will be seen when a particular option of the facet is selected, the eCommerce Search system should properly understand the complex hierarchical data structures that are used to manage product information. If a product is displayed in search results, the customer should be confident that the combination of product features indicated by their search phrase and their facet selection are indeed available on the same purchasable item.

Grid Dynamics

Besides intrinsic product characteristics like brand, model, and size, there are a variety of other essential characteristics for an excellent search experience like per store inventory, per contact channel / promotion campaign / shopper location, price, and product availability. Search results must contain only items satisfying business rules and constraints.

Traditional indexing does not scale well when dealing with product groups or collections nor with products that have large numbers of variants that are different for example only in size and color. Grid Dynamics has extensive experience developing enhanced indexing procedures to support products with complex product relationships which allows for both scale and efficiency.

Customer keyword search requests can match on attributes of a product or with some of its variants (or skus). Similarly facets or filters applied to search results can narrow down the result set at the product or variant level. Accurately and efficiently handling these distinctions requires a highly configurable indexing engine. The benefit to the customer is that she can be confident that the products being shown satisfy all of the preferences she has expressed through the words in her search phrase and the faceted refinements she has selected.

Support of product collections is a challenge for many search engines. Products can belong to multiple collections creating complex many-to-many relationships which are difficult to manage and search over. In various scenarios, merchandisers want search results to show only collections, or only products belonging to collections, or even mixed products and collections. Grid Dynamics knows how to support all those complex collection search scenarios.

eCommerce navigation involves reading static product data, but up-to-date items availability helps to avoid over-selling products. Traditional inverted index based search engines are poorly suited for frequent updates. Grid Dynamics has created special support for near-real time updates of availability information which guarantees that customers are always presented with products they can actually purchase.

customer

eCommerce search engines have to "understand" the customer, building meaningful hypotheses on what exactly a customer means by her search phrase, supporting domain-specific linguistics, normalizations, and spelling corrections. The search engine has to find a perfect balance between search result precision and

recall for each customer query. This "Concept Search" goes far beyond simple keyword matching and allows the search engine to precisely understand customer need and, in many cases, to present the customer with a small set of products which perfectly match the need.

Classical text matching algorithms work poorly for eCommerce. High recall queries tend to create a mess in search results and return thousands of weakly relevant products which pollute the faceted experience. To address this issue, Grid Dynamics devised multi-stage matching which allows the search engine to find the right balance between precision and recall and eliminate unwanted weakly relevant facet values.

Classical relevancy scoring models, such as idf-tf, are not the best performers for eCommerce, as they are designed for unstructured long text documents but not for relatively brief yet well-structured product documents. Grid Dynamics created a scoring algorithm which is optimized for small, structured documents and takes into account dozens of signals about product relevancy to particular request in particular context and channel.

merchandiser

eCommerce search engines should provide merchandisers with simple yet powerful tools to promote particular products and finding the best match between what the customer wants to buy and what Merchandiser wants to sell. Grid Dynamics has built a flexible rules engine which allows merchandisers to tightly control almost every aspect of search experience.

Grid Dynamics provides the ability to preview and explain search results on the level which can be well understood by non-technical people. Merchandisers are able to tweak matching configurations to achieve the best possible experience.

Merchandisers like to experiment with business rules and linguistics and see how the system behaves with new data. Grid Dynamics can help update the linguistics and business rules at near real time, giving every opportunity for experimentation.

It is critical for merchandisers to understand how well the search system performs, what customers are looking for and how efficient search is for them. Grid Dynamics leverages an automated search optimization approach which continuously analyses search quality and allows merchandisers to monitor the behavior of the full range of search phrases used by site visitors, identifying product attribution linguistics and inventory gaps.

stability and elasticity

Support of the full enterprise catalog of a major brick-and-mortar retailer requires a search engine to support massive clustering and sharding. High performance of a search engine is a prerequisite to be able to support many complex business use cases which may require multiple roundtrips to the search engine per single user request. In eCommerce, a successful marketing campaign or sales event can cause sharp spikes in the traffic, 5-10x of the normal day request volume. Search engines need to successfully respond to queries even in resource congestion situations. For major sales events, such as Black Friday or Cyber Monday, it should be easy to roll out new service instances depending on expected spikes in user traffic.

Grid Dynamics is able to gracefully degrade search experience in resource congestion situations to trade search quality for performance and always stay up. High performance index replication infrastructure allows to easily grow the capacity by adding new search servers.

Solr is the most extensible search platform for eCommerce sites but there is a lot of technical effort that goes into configuring it for your business. Grid Dynamics has the Solr expertise to get you there and to help you maintain it. Call Grid Dynamics today for a search assessment, replatform, or search tuning. Solr services is just one of the many offerings at Grid Dynamics. To learn more about how we can help accelerate your business, contact us at sales@griddynamics.com.

about Grid Dynamics

Grid Dynamics is a leading provider of open, scalable, next-generation commerce technology solutions for Tier 1 retail. With in-depth expertise in commerce technologies, wide involvement in the open source community and a modern, global workforce, Grid Dynamics helps great companies gain a sustainable business advantage by implementing and managing solutions in the areas of omnichannel platforms, product search, and continuous delivery. To learn more about Grid Dynamics, find us at www.griddynamics.com or by following us on Twitter @GridDynamics.

Grid Dynamics