

d u l a n c e

● ● ● ● ● ● ● ● ● ● ● ● ● ● ● PRODUCT SEARCH

The “Long Tail” shopping search

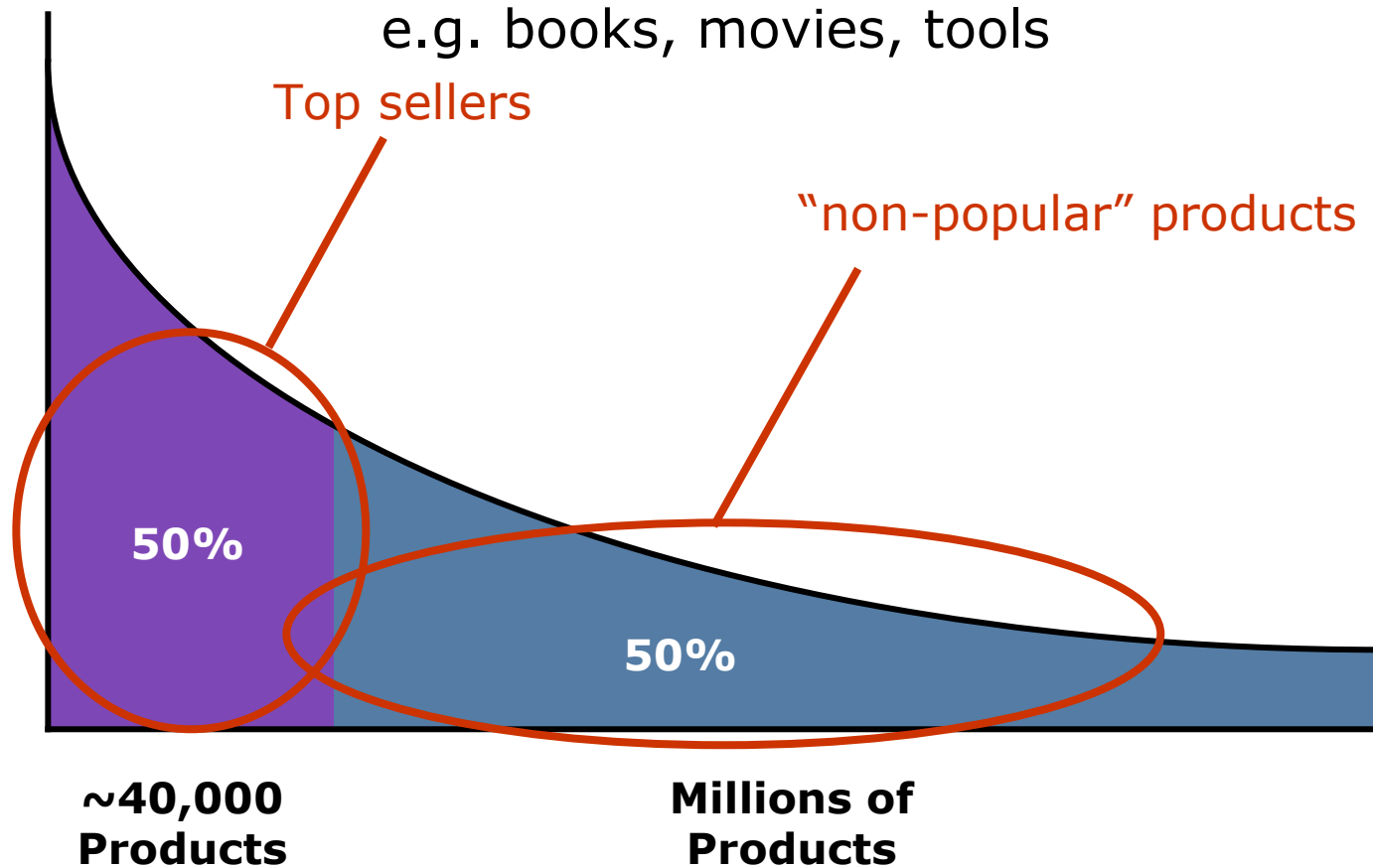
Shopping Search Evolution



- Still no shopping solution for the entire Web
- Web search: too broad and delayed
 - Too many irrelevant results (90%)
 - No prices, PageRank doesn't work for shopping, etc.
 - Often obsolete data (6 weeks re-crawl cycle)
- 1st generation comparison shopping: too narrow
 - Limited choice of products and vendors
 - Typically only 20,000 out of ~1 million online stores
- 2nd generation: Dulance 1.0 and Froogle
 - All the web, but still delayed
- 3rd generation: Dulance 2.0 (RSS or WS)
 - All the web and "Search the Future"

The “Long Tail”

Sales for any given product group
e.g. books, movies, tools



Dulance 1.0: Long Tail



- Algorithmic approach to shopping search
- Scope: all online stores
 - Over 1 million stores in N. America
- Filter that recognizes online stores
 - Repository of all e-commerce pages on the web
- Extracts product name, price, used/new, etc.
- Doesn't require site-by-site screen scraping
- New page relevancy ranking algorithm
- Price cluster analysis

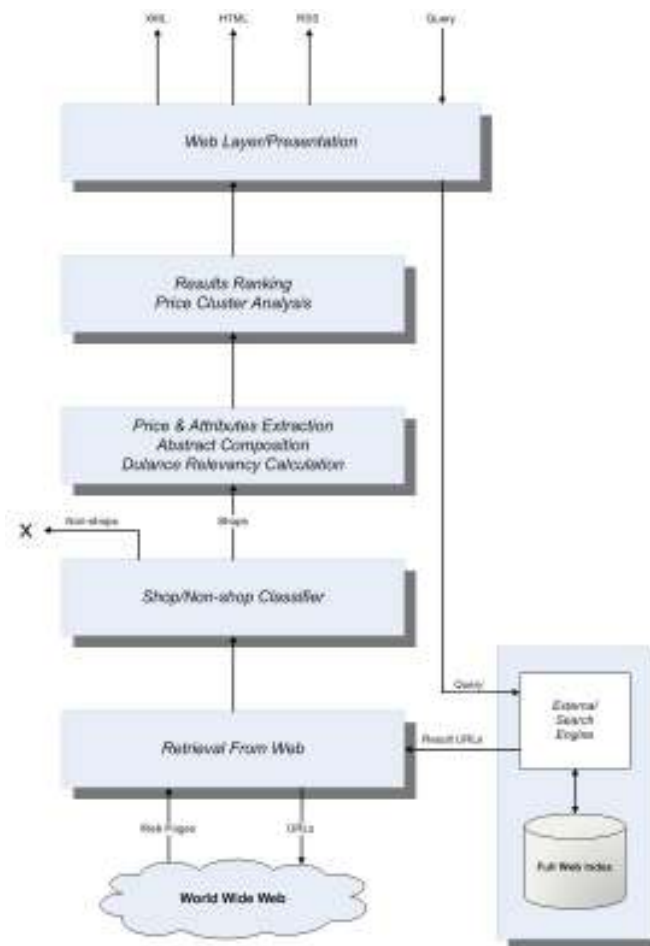
Benchmark Searches



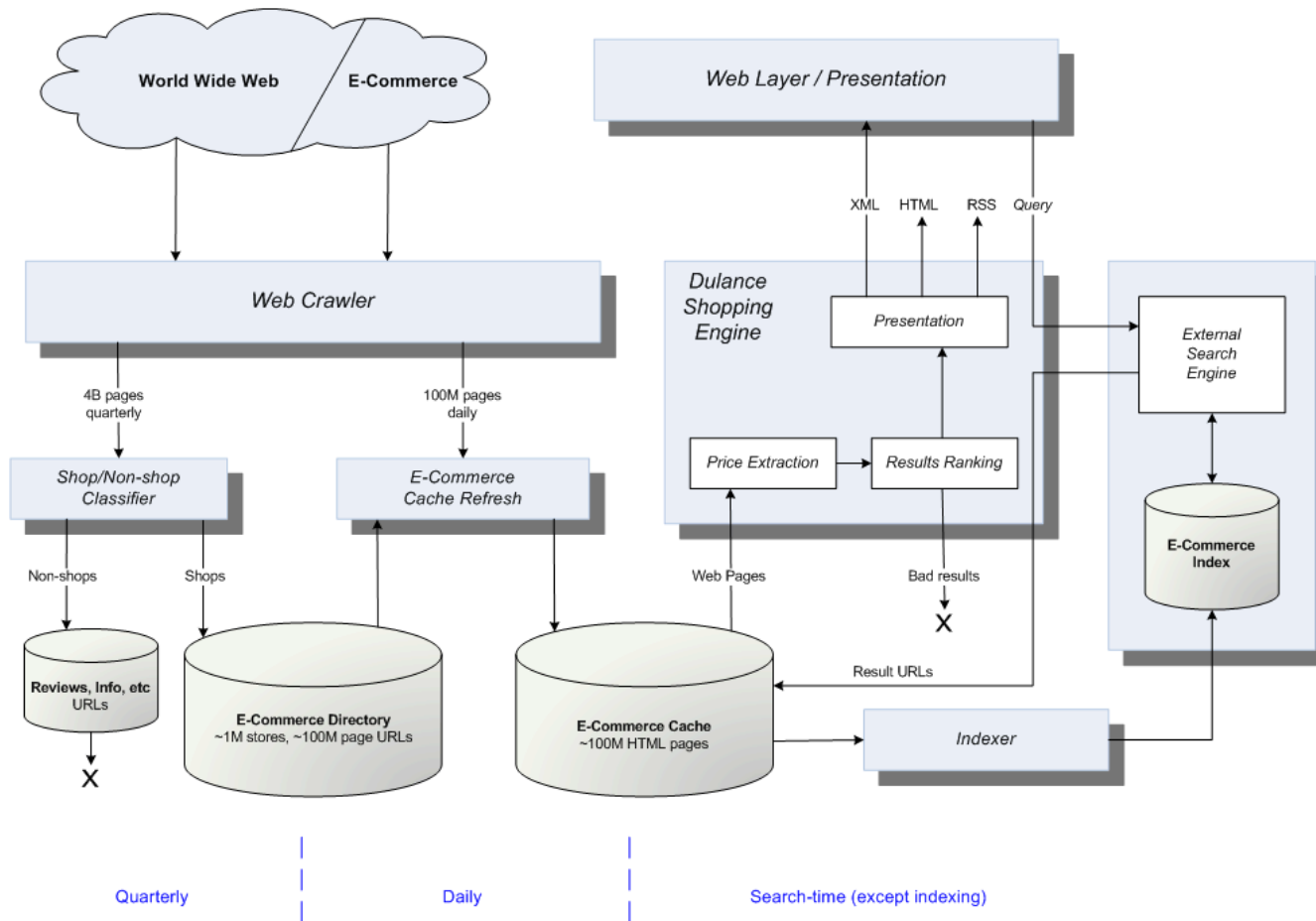
Number of product results found

Query	Dulance	Shopzilla
Fischer Worldcup RC	9	1
Fanatic Goya	7	0
Ezzy Wave 5.0	9	0
Z-coil Freedom 2000	6	0
Magma Kohntarkosz	20	1
Martin 000-28EC	30+	1
Fortigate 60	30+	1
185 Sullair air compressor	9	0
EH-76 vertical reciprocating conveyor	3	0
Keil PK51	4	: -)

No-Cache Architecture



Architecture Overview



Page Classifiers

- Flexible technologies with a range of applications
 - Shop/non-shop filter
 - Used/new filter (for eBay)
 - Can be applied to travel, jobs, people, restaurants, etc.
- Linear supervised statistical machine learning
 - Attribute reduction
 - Hand-picked and/or automatically selected attributes
 - Expert rules (synonyms) into attributes
- Optimized for webpages (HTML, JavaScript)
- Query-independent; crawl-time
- Unique value:
 - New approach to categorization
 - Work-years tuning the models

Relevancy

- Optimized for product search / shopping
- Is the product sold here?
 - Boolean relevancy: Yes/No
- Query-dependent; search-time
 - No link analysis required
- HTML-aware proximity measure
 - Measure how dense query words are
 - Add premium/penalty for Title, H1, , hyperlinks, etc.
 - Combine these to calculate relevancy
 - If >70% then true

Attribute Extraction

- On-the-fly extraction, no normalized product DB
 - Query-dependent; search-time
 - Pre-parsing for performance
- Probabilistic approach – no expert rules
- Price, availability, etc.
 - HTML-aware (tags, CSS, tables, etc.)
 - Disqualify S/H, MSRP, “you save”, ...
 - Disqualify accessories, related items, ...

Other Components

- Price cluster analysis
- Accessories filter
- Affiliates, spam, etc. filters
- Address grabber
 - Finds "Contact us" webpages, grabs addresses

Additional Applications



- RUsure Browser Toolbar
 - Automatically initiates price comparison when the user is at eBay, Amazon, etc.
 - <https://addons.update.mozilla.org/extensions/moreinfo.php?id=652>
- Local shopping
 - Find specific products sold in local stores
 - Requires much larger pool of merchants than Shopzillas have
- Blending shopping into web search
 - Enhance generic web search results with extracted prices and availability
 - No shopping tab required
- SpamFree: the Web without shops
- Contextual advertising
 - From product reviews to instant price quotes, in one click
 - <http://dulance.com/tripleclick.html>

Search the Future



- Shoppers register their wish-lists and receive notifications when prices drop
 - via RSS or e-mail
- Merchants publish time-sensitive offerings and promotions and have them delivered, nearly instantly, to the awaiting shoppers
 - Empower sellers to *sell*
 - Matching buyers' wish lists
 - Instant index update, no delays

Company Overview



- Sergei Burkov, Ph.D. - Founder & CEO
 - Serial entrepreneur: Bilbo Innovations, Invincible: data security company, sold to VASCO (NASDAQ: VDSI)
- Igor Gusak, Ph.D. –VP Business Development
 - VP of Sales and GM of Commtouch (NASDAQ: CTCH)
- 8 highly skilled mathematicians and software developers in Moscow, Russia
 - 2 Ph.D.s; 5 from Math Dept of Moscow Univ.; 2 of MIPT
- 2+ years of development and customer testing

Revenue Opportunity



Monetization: sponsored listings

Premium stores

50,000 stores

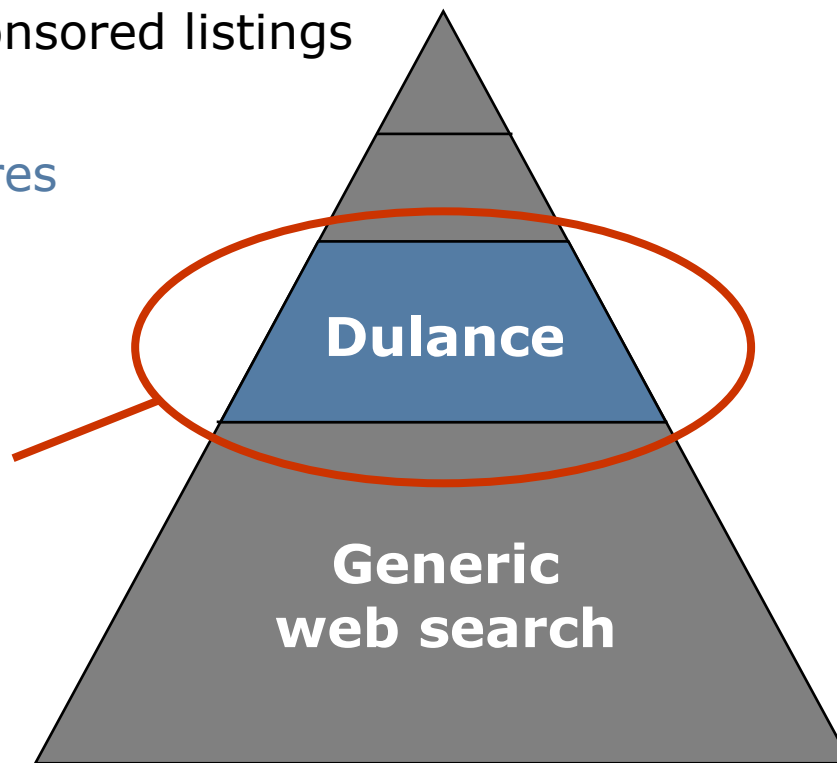
All stores

1 million stores

New opportunity

Non-stores

30 million sites



$$\text{Revenue} = \text{Traffic} * \$30 \text{ CPM}$$

Dulance Value Proposition



- Value to consumers
 - Find any product, from any merchant, at lower prices
 - Shopping agent that watches for *future* sales and promotions
- Value to merchants
 - More qualified leads - higher ROI
 - Push limited-time promotions to consumers