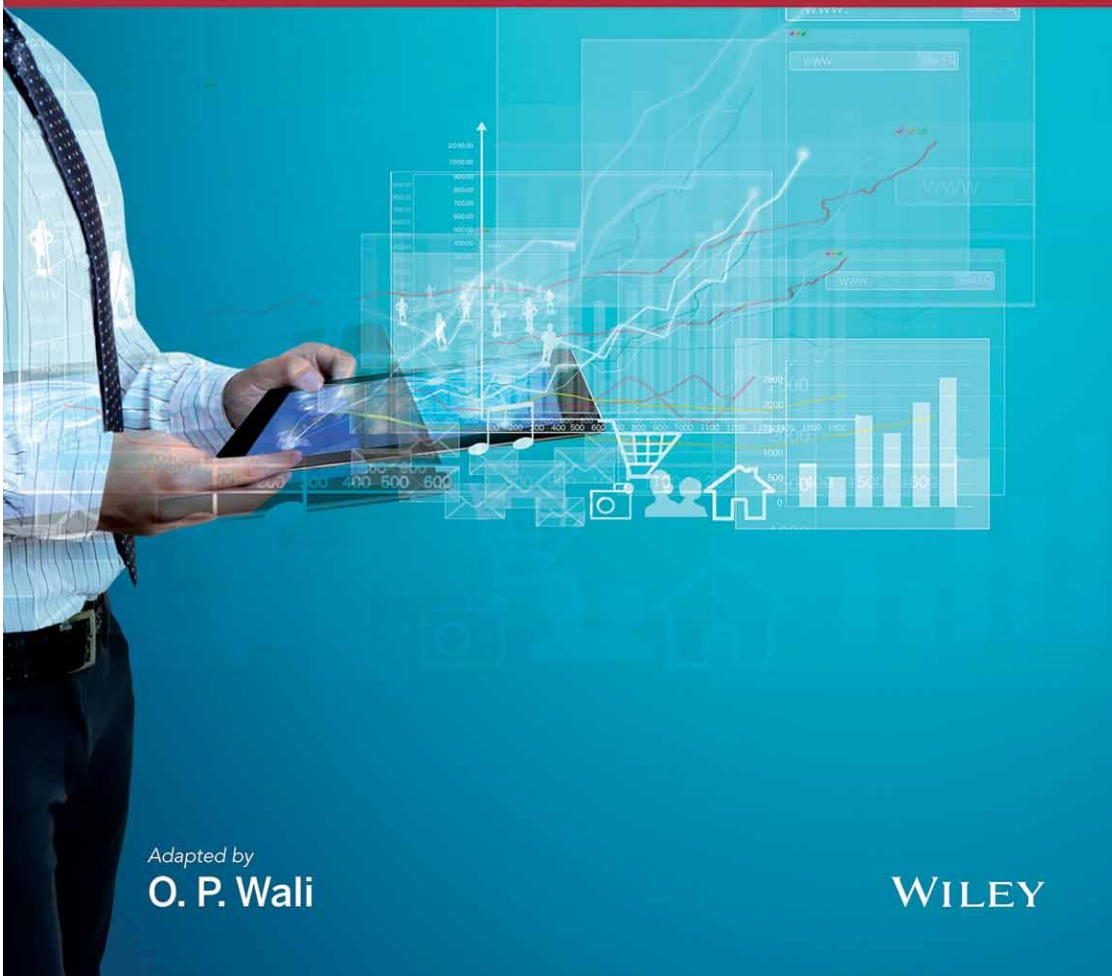


TURBAN | VOLONINO | WOOD

Information Technology for Management

Advancing Sustainable, Profitable Business Growth



Adapted by
O. P. Wali

WILEY

ns, Inc.

Chapter 6

Internet Technologies and Search Strategies

Chapter Outline

1. [Search Technology](#)
2. [Organic Search and SEO](#)
3. [Paid Search Strategies and Metrics](#)
4. [Semantic Web Search](#)
5. [Recommendation Engines](#)

1. Search Technology

Why search is important for Business?

Google Data Centers

Not only does Google maintain a copy of the Internet for its search engine services, it is also constantly updating a map of the entire planet for users of its popular Google Earth application. In addition, the company is making a full-text, searchable copy of all 129,864,880 known books, equal to 4 billion pages or 2 trillion words. And then there are applications like Gmail, serving roughly 425 million people and YouTube, where 100 hours of video are uploaded every minute! Add all this up, and Google is facing perhaps the biggest data storage challenge ever. So where does Google store all of these data?

Google Data Centers

Challenges: Energy, Performance, and Security

Information collected by Google is housed on over 1 million servers spread across 12 different facilities worldwide. The facilities are large, factorylike installations containing row upon row of racked and stacked servers. Cooling systems, required to keep servers from overheating, are a significant component of any large data center (Figure 6.7). Google pioneered the software systems that connect the company's servers and make it possible for various applications to access data stored on the machines. Unlike other companies that purchase servers from outside suppliers, Google builds its own. Based on its experience creating the hardware, software, and facilities necessary to power the company on a global scale, Google is recognized as a leader in data center operations.



Bloomberg/Getty Images

Figure 6.7 Pipes pass through the chiller plant at the Google, Inc. data center in Changhua, Taiwan. Google doubled its spending plan for its new data center in Taiwan to \$600 million amid surging demand from Asia for its Gmail and YouTube services.

Google Data Centers

The company's data centers, including the servers, are built with energy efficiency, reliability, and performance in mind. As Google is a leading provider of Internet services, its data infrastructure must keep up with growing consumer demand for speedy performance and reliability. A typical Google search delivers millions of pages of results in less than half a second. Consumer expectations for performance have grown so high that waiting more than a few seconds for an e-mail to load or a search to run can cause frustration.

Google Data Centers

More recently, Google has had to contend with revelations that the U.S. National Security Agency (NSA) breached its server network security. This follows cyberattacks in 2010 and 2011 by hackers suspected of being associated with the Chinese government. Protecting company data from criminals is a significant challenge in itself, but Google is understandably frustrated by the fact that it must now fight off attacks from two world superpowers, one of which is its own government.

Environmental Impact

Industrywide, data centers used 38,000 megawatts of electricity in 2012, representing a tremendous increase of over 63 percent from the previous year. Industrywide, data center energy use and the related environmental impact have become an issue of growing concern. Google is widely recognized as operating some of the most efficient data centers in the world, but many critics are disturbed by the industry's overall level of energy consumption. According to some estimates, data centers account for about 2 percent of the world's energy use and the fast rate of growth is cause for concern (see Figure 6.8). Google has taken an active approach to reducing its environmental footprint by increasing the company's use of renewable energy sources and continuously developing new ways to become more energy-efficient. (For further information, see google.com/about/datacenters.)

Google Data Centers

Google Data Center Statistics

- **Number of servers worldwide:** Over 1 million
- **Number of data centers:** 6 to 12 in North America, 1 in South America, 2 in Asia, and 3 in Europe
- **2013 capital investment in data centers:** \$7.35 billion
- **Estimated investment in data centers since 2006:** \$21 billion
- **Data processing volume:** Over 20 petabytes a day
- **Average energy efficiency:** PUE* = 1.12
- **Energy use:** Continual use of about 260 megawatts of electricity, approximately 0.01 percent of global energy consumption
- **Energy use comparisons:** Owns about 3 percent of servers worldwide, but only uses about 1 percent of data center industry energy
- **Renewable energy:** Claims that approximately 35 percent of its energy use comes from renewable sources

Google Data Centers

*PUE stands for **Power Usage Effectiveness**. A PUE of 2.0 means that for every watt of power devoted to computing, an additional watt is spent on cooling, power distribution, and overhead. The Data Center Industry average PUE falls between 1.8 and 1.89.

Sources: Anthony (2013), Miller (2013), Glanz (2011, 2012), Venkatraman (2012), Schneider (2011), Gallagher (2012), Grifantini (2011), Newman (2011), Jacobson (2010).

Using Search Technology for Business Success

- Why Search is Important for Business
 - Enterprise search tools allow organizations to share information internally.
 - An organizations' ability to share knowledge among employees is vital to its ability to compete.
 - Information is not always in the same format.

Using Search Technology for Business Success

- Why Search is Important for Business
 - Structured data: information with a high degree of organization, such that inclusion in a relational database is seamless and readily searchable by simple, straightforward search engine algorithms or other search operations.
 - Unstructured data: “messy data” not organized in a systematic or predefined way.

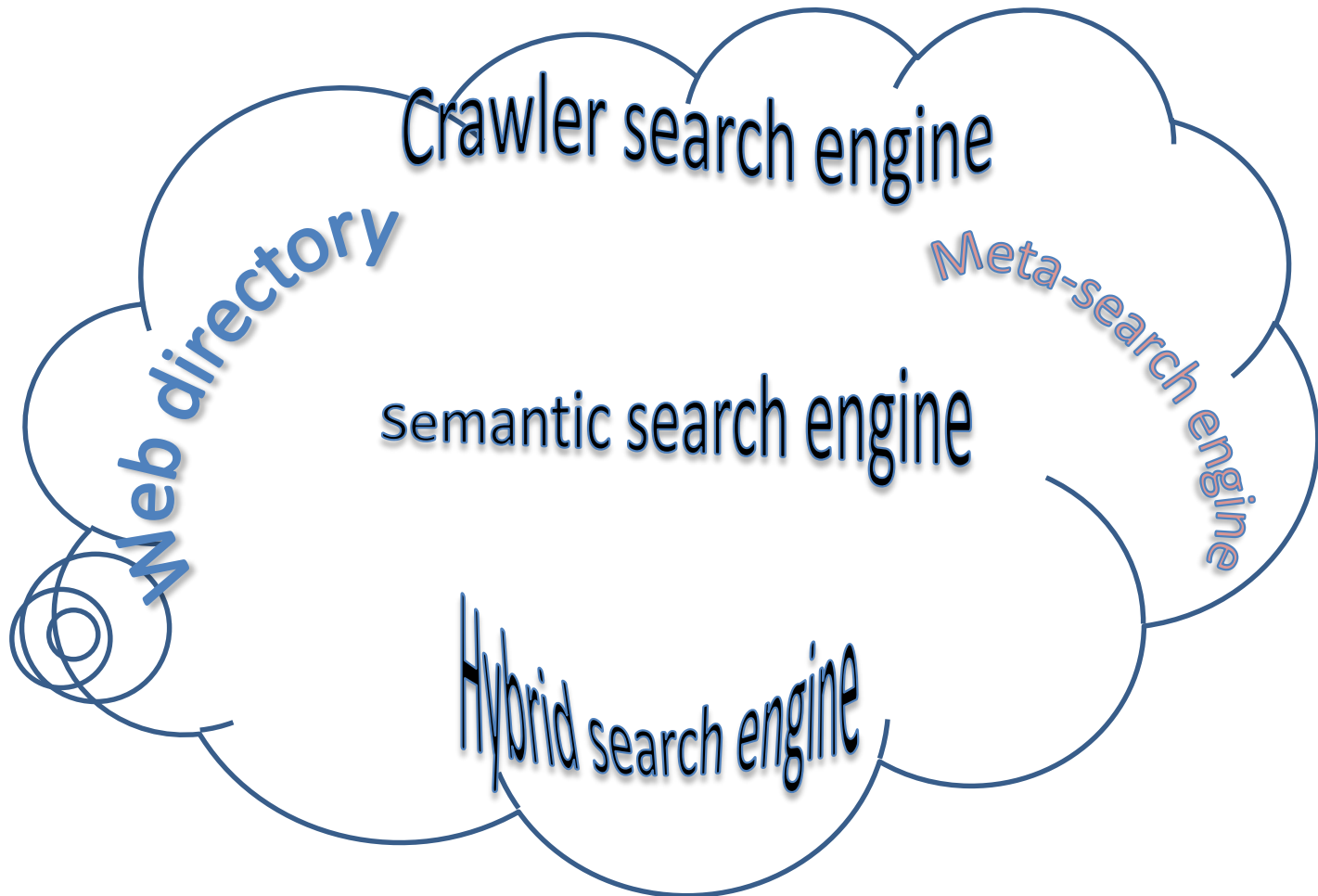
Using Search Technology for Business Success

- How Search Engines Work
 - Search Engine: an application for locating webpages or other content on a computer network using **spiders**.
 - Spiders: web bots (or bots); small computer programs designed to perform automated, repetitive tasks over the Internet.
 - Bots scan webpages and return information to be stored in a page repository.

Using Search Technology for Business Success

- Web Directories
 - Typically organized by categories.
 - Webpage content is usually reviewed by directory editors prior to listing.
 - Page Repository: data structure that stores and manages information from a large number of webpages, providing a fast and efficient means for accessing and analyzing the information at a later time.
 - Dmoz.org, botw.org, looksmart.com, business.com

Using Search Technology for Business Success



Using Search Technology for Business Success

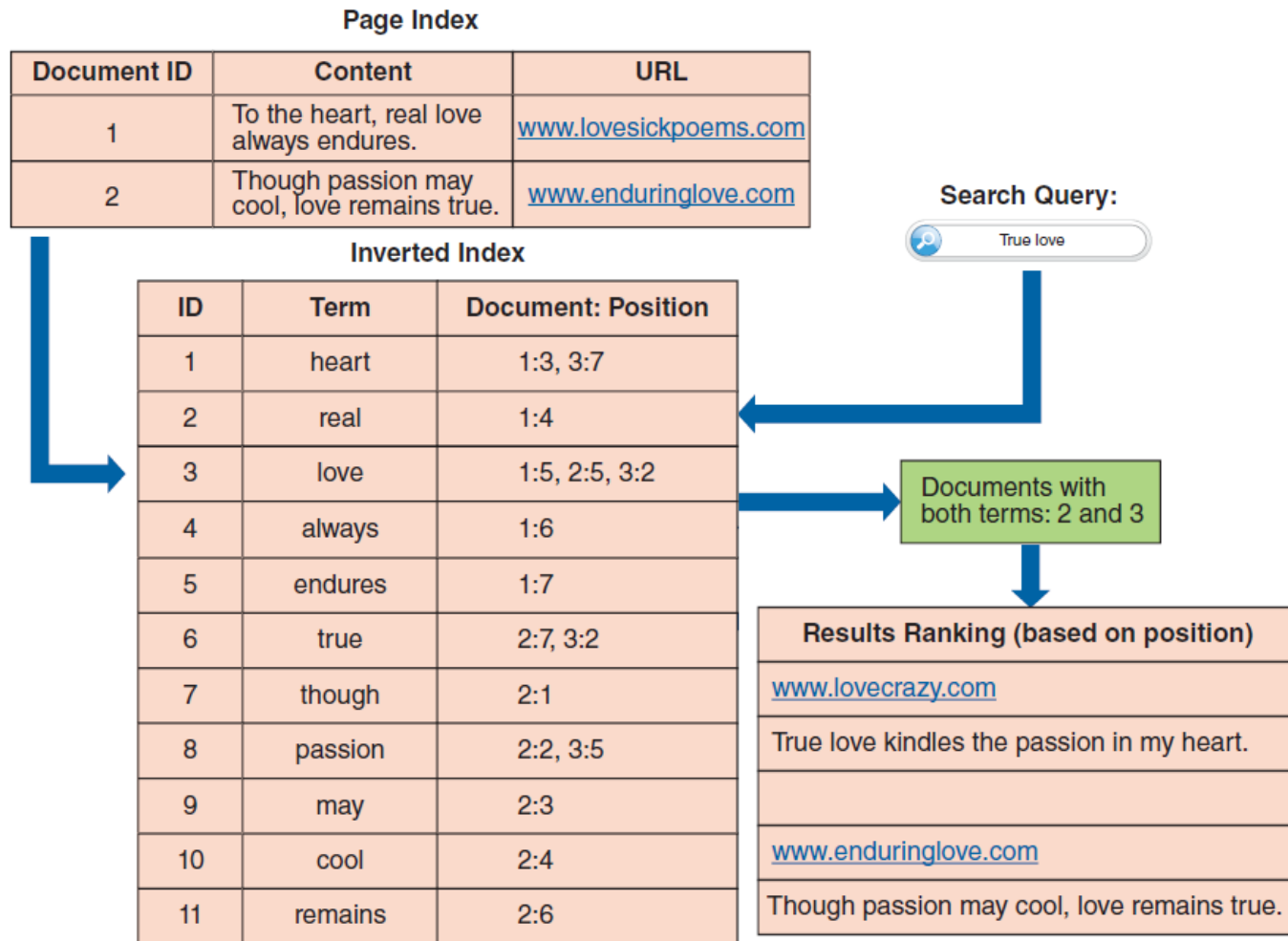


Figure 6.6 Search engines use inverted indexes to efficiently locate Web content based on search query terms.

Using Search Technology for Business Success

- Security Issues
 - Limited access to certain data via job function or clearance.
 - Request log audits should be conducted regularly for patterns or inconsistencies.

Using Search Technology for Business Success

- **Enterprise Search** is used to describe the software of search information within an enterprise (though the search function and its results may still be public).
- Enterprise search can be contrasted with web search, which applies search technology to documents on the open web, and desktop search, which applies search technology to the content on a single computer.

Using Search Technology for Business Success

- Enterprise Vendors
 - Specialized search vendors: Software designed to target specific user information needs
 - Integrated search vendors : Software designed to combine search capabilities with information management tools
 - Detached search vendors : Software designed to target flexibility and ease of use

Using Search Technology for Business Success

- Recommendation Engines
 - Attempt to anticipate information users might be interested in to recommend new products, articles, videos, etc.

Using Search Technology for Business Success

- Search Engine Marketing
 - A collection of online marketing strategies and tactics that promote brands by increasing their visibility in search engine results pages (SERPs) through optimization and advertising.



Using Search Technology for Business Success

- Search Engine Marketing
 - Basic search types:
 - Informational search
 - Navigational search – to find website
 - Transactional search
 - Strategies and tactics produce two outcomes:
 - Organic search listings
 - Paid search listings
 - Pay-per-click (produce click-through rates)
 - Social media optimization

Using Search Technology for Business Success

- Mobile Search
 - Technically configured mobile sites
 - Content designed for mobile devices
 - Showrooming – mobile shoppers use barcode scanning or QR code scanning for information gathering and transacting

TABLE 6.2 How In-Store Shoppers are Use Mobile Devices

Shopping Behavior	Percentage
Use a mobile device to assist with shopping at least once a month	62%
Use a mobile device to assist with shopping at least weekly	17%
Use their smartphones for more than 15 minutes while in a store	~50%
Use a search engine to browse product information while shopping in a store (Search engines dominate store websites, brand websites, store apps, and brand apps as a starting point.)	82%
Use their smartphone for preshopping activities such as finding store locations, finding product reviews, making price comparisons, finding promotional offers, and so on	90%

Top mobile shopping categories include:

Appliances	97%
Groceries	89%
Baby care	87%
Electronics	87%
Household care	86%

Across all product categories, mobile shoppers report using their smartphones instead of asking for information from store employees	33%
--	-----

Use smartphones instead of asking for information from store employees when shopping for appliances	55%
---	-----

Use smartphones instead of asking for information from store employees when shopping for electronics	48%
--	-----

In-store price comparisons are the most common shopping activity across product categories

Appliances	74%
Electronics	70%
Baby care	62%
Household care	58%

Mobile shoppers prefer to view mobile websites vs. mobile apps	65% vs. 35%
--	-------------

Source: ThinkWithGoogle.com (2013).

2. Organic Search and Search Engine Optimization

Organic Search and Search Engine Optimization

- Search Engine Optimization
 - Keyword conversion rates: the likelihood that using a particular keyword to optimize a page will result in conversions*.
 - Ranking factors
 - Reputation or popularity
 - PageRank: Google's algorithm based on the assumption that people are more likely to link a high-quality website than poor-quality site.
 - Backlinks: external links that point back to a site.
 - Relevancy
 - User Satisfaction

Conversions: when a website visitor converts to a buyer

Popularity Factors used by SEs

TABLE 6.3 Popularity Factors Used by Search Engines

Social media indicators: Webpages being discussed in social media are more likely to be popular. Examples include comments on Facebook and Google+, shares, Tweets, Likes, and so on.

Site traffic: The more traffic a site gets, the more likely it is to contain quality content.

Listings in quality Web directories: Sites that have been accepted for listing in quality Web directories are more likely to contain quality content because they have been reviewed by human editors.

Reputation on review sites: Sites with high ratings on review sites (such as Yelp.com, Zagat.com, and Epinions.com) are more likely to have quality content.

PageRank of sites containing backlinks: Backlinks from sites with high PageRank scores themselves are thought to be more valuable.

SERP click-through rate (CTR): Google and other search engines monitor how often searchers click through on a SERP listing. People tend to click listings that appear to contain quality content.

Onpage factors: Metadata (e.g., page titles, page descriptions) and descriptive URLs influence how a listing appears on a SERP. This, in turn, can affect the CTR.

Dwell time and bounce rate: **Dwell time** is the amount of time a user stays on a page before returning to the SERP. Longer dwell times suggest users found content that was worth reading. A “bounce” is an extremely short dwell time—in other words, a user who clicks on a listing and comes right back is said to “bounce.” Low **bounce rates** and long dwell times correlate with high content quality.

Factors affecting Relevance Score

TABLE 6.4 Factors That Affect Relevance Scores

Keywords related to the search topic suggest relevant content.

Page titles: Words in the page title that are related to the topic suggest relevant content.

Relevant phrases in text: In addition to keywords, search engines look at the words and phrases on the page to determine relevance.

Amount of text on page that appears relevant: The proportion of relevant text to nonrelevant text can influence relevance.

Backlinks from relevant sites and Web directories: Webpages that are listed in relevant categories of Web directories are more likely to be relevant because they were reviewed by human editors.

SERP click-through rate (CTR): Searchers are more likely to click on listings that contain relevant content.

Onpage factor: Metadata (such as page titles, page descriptions) and descriptive URLs should reflect the page content. People use the information in search listings to determine if a link contains relevant information. This affects CTR.

Dwell time and bounce rate are impacted by how relevant a website's content is. Long dwell times and short bounce rates suggest relevant content related to the search.

Factors affecting User Satisfaction

TABLE 6.5 Factors Affecting User Satisfaction

Dwell time: Users who stay on a site longer are probably more satisfied.

Site speed: Slow page loading time on websites reduces satisfaction.

Reading level: Reading levels that are too high or too low frustrate users.

Hacked sites, malware, spam reduce user satisfaction significantly.

Website satisfaction surveys: Google created user satisfaction surveys that webmasters can embed in their websites. Positive responses to these surveys can improve ranking.

Barriers to content: Making people register, provide names, or fill out forms to get to content has a negative impact on user satisfaction.

Other factors: Too many ads, page-not-found errors, duplicate content/pages, content copied from other websites, and spam in comment sections all detract from user satisfaction.

Organic Search and Search Engine Optimization

- Inbound marketing
 - An approach to marketing that emphasizes SEO, content Marketing, and social media strategies to attract customers.
- Outbound marketing
 - Traditional approach using mass media advertising.

Inbound Outbound Marketing

Inbound Marketing

Communication is interactive or two-way

Customers seek out the business

Marketers promote company by educating or entertaining

Typical strategies: content, social media & SEO

Outbound Marketing

Communication is one-way

Businesses seek out customers

Businesses broadcast messages that interrupt customers

Typical strategies: TV, radio, print, outdoor, cold calls

Organic Search and Search Engine Optimization

- Black Hat SEO
 - Gaming the system or tricking search engines into ranking a site higher than its content deserves.
 - 1. Link spamming: generating backlinks toward SEO, not adding user value.
 - 2. Keyword tricks: embedded high-value keywords to drive up traffic statistics.
 - 3. Ghost text: text hidden in the background that will affect page ranking
 - 4. Shadow (ghost or cloaked) pages: created pages optimized to attract lots of people through redirect.

3. Paid Search Strategies and Metrics

Pay-Per-Click and Paid Search Strategies

- Pay-Per-Click
 - PPC advertising campaigns:
 1. Set an overall budget
 2. Create ads
 3. Select associated keywords
 4. Set up billing account information
- Keywords and Bid Pricing
- Geographic Location and Time of the Day
- Landing Pages

Pay-Per-Click and Paid Search Strategies

- Paid Search Advertising Metrics
 - Click through rates (CTR): used to evaluate keyword selection and ad copy campaign decisions.
 - Keyword conversion: should lead to *sales*, not just visits.
 - Cost of customer acquisition (CoCA): amount of money spent to attract a *paying* customer.
 - Return on advertising spend (ROAS): overall financial effectiveness.

4. Semantic Web and Search

A Search for Meaning—Semantic Technology

- Semantic Web
 - Meaningful computing using metadata: application of natural language processing (NLP) to support information retrieval, analytics, and data-integration that compass both numerical and “unstructured” information.
- Semantic Search
 - Process of typing something into a search engine and getting more results than just those that feature the exact keyword typed into the search box.
- Metadata
 - Data that describes and provides information about other data.

A Search for Meaning—Semantic Technology

TABLE 6.6 Evolution of the Web

Web 1.0 (The Initial Web) A Web of Pages	Pages or documents are “hyperlinked,” making it easier than ever before to access connected information.
Web 2.0 (The Social Web) A Web of Applications	Applications are created that allow people to easily create, share, and organize information.
Web 3.0 (The Semantic Web) A Web of Data	Information within documents or pages is tagged with metadata, allowing users to access specific information across platforms, regardless of the original structure of the file, page, or document that contains it. It turns the Web into one giant database.

A Search for Meaning—Semantic Technology

- Web 3.0
 - Developed by W3C – World Wide Web Consortium
 - Resource description framework (RDF)
 - Used to represent information about resources
 - Web ontology language (OWL)
 - Language used to categorize and accurately identify the nature of Internet things
 - SPARQL protocol
 - Used to write programs that can retrieve and manipulate data stored in RDF
 - RDF query language (SPARQL)

A Search for Meaning—Semantic Technology

- Semantic Search Features and Benefits
 - Related searches/queries
 - Reference results
 - Semantically annotated results
 - Full-text similarity search
 - Search on semantic/syntactic annotations

5. Recommendation Engines

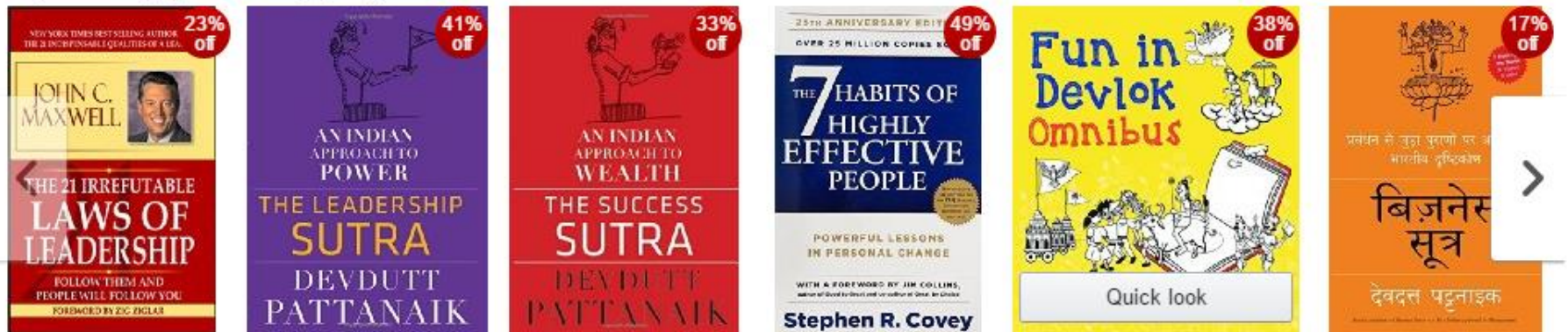
Recommendation Engines

- Recommendation Filters
 - Content-based filtering: products based on product features in past interactions.
 - Collaborative filtering: based on user's similarity to other people.

Recommendation Engines



Inspired by your shopping trends



Related to items you've viewed [See more](#)



Recommendation Engines

- Limitations of Recommendation Engines
 - Cold start or new user: challenging since no starting point or preexisting information exists.
 - Sparsity: unable to create critical mass due to few ratings or similar groups are unidentifiable.
 - Limited feature content: manual information entry is prohibitive where there are many products.
 - Overspecialization: narrowly configured results may only recommend the same item, but in different sizes or colors.

Recommendation Engines

- Hybrid Recommendation Engines
 - Weighted hybrid: results from different recommenders are assigned weight and combined numerically to determine final recommendations.
 - Cascade hybrid: results from different recommenders assigned a rank or priority.
 - Mixed hybrid:
 - Results from different recommenders presented alongside of each other.
 - Results from different recommenders combine results from two recommender systems from the same technique category.

TABLE 6.7 Examples of Recommendation Engine Applications

Company	How It Uses Recommendation Engines. . .
Amazon	Recommends products using multiple filtering methods.
Netflix	Approximately 75 percent of Netflix movies are selected as a result of its recommendation system.
Pandora	This streaming-music site creates playlists based on similarity to initial songs or artists selected by the user.
CNN, Time, Fast Company, Rolling Stone, NBCNews.com, Reuters, Us Weekly	These news and entertainment companies all use a recommendation engine (or “content discovery system”) created by Outbrain.com to suggest additional articles related to the one site visitors initially viewed.

YouTube	YouTube uses a variation of Amazon's recommendation engine to suggest additional videos people might like to watch (Linden, 2011).
Goodreads	This social website for readers recommends books based on user ratings of books they have read.
Samsung	Uses recommendation engines built into its "smart TVs" to suggest television programming to viewers.
Facebook and LinkedIn	These social networking services use recommendation engines to suggest people that users may want to connect with.
Apple	Helps users find mobile apps they might enjoy.
Microsoft Xbox 360	Suggests new games based on what users have previously shown an interest in.
Tripadvisor	Recommends travel destinations and services based on destinations people have viewed or rated.

Stitch Fix

This fashion start-up uses a recommender system in conjunction with human stylists to select and ship clothing products to customers, before customers viewed or ordered them!

Fashioning Change

This online clothing retailer lets customers select what their recommendations are based on: style preferences, personality, causes, and “likes.”