

Web and Social Media Analytics

Search engines – part II

Dr Philip Worrall

School of Computer Science and Engineering
115 New Cavendish Street
University of Westminster
London, W1W 6UW
worrallph@westminster.ac.uk

LW4

With acknowledgement and thanks to Vassiliki Bouki

Recap: Search engines and IR systems

1. The web has grown exponentially since its inception and roughly **doubles** in size every 3 years.
2. Search engines provide a solution to **INFORMATION OVERLOAD**, this would otherwise make finding relevant information on the web inefficient and time consuming.
3. Significant progress was made in the early 2000s with the application of **information retrieval theory** to the problem of internet search.
4. Modern search engines use complex **retrieval models** to identify relevant content on the web and determine an appropriate ordering.
5. PageRank™, proposed by Google, is an example of an algorithm used to **objectively quantify** the quality of a web page based on the number of inbound links.

Week 4

Plan for today

- ❖ Revisit Ash's Marketing Funnel
- ❖ Search engine optimisation (SEO) strategies
- ❖ Search engine marketing (SEM)
 - ❖ Pay-per-click (PPC)

What is SEO?

Search Engine Optimisation (SEO): ‘Once highly *specialised task* relegated to the backroom of a website development team is now a *mainstream marketing activity*’ (from ‘The Art of SEO’)

SEO = how to let search engines (and audience) know about your website

SEO Goals:

- I. Increase visibility of the site
- II. Increase relevant traffic to the site
- III. Increase return on investment (ROI) – SEO works hand-in-hand with Analytics

Search results: Organic vs. Paid

There are two kinds of results you get from a search engine:
‘Organic’ and **‘Paid’** results

‘Organic’ results: results produced ‘directly’ by a search engine – the web page listings that are produced by a search engine and are designed to closely match a user’s query.

‘Paid’ results → advertisements

Why should you want to ***rank first*** in ‘organic’ results (but not necessarily in ‘paid’ results)?

Example. Bing.com

Paid results



Microsoft Bing

Search: eye glasses

ALL SCHOOL SHOPPING IMAGES VIDEOS MAPS CHAT MORE

About 2,530,000,000 results Date Results near London, London Change

Opulize Pop 4 Pack Retro Round Black Bl... £10.48 Amazon UK	The Reading Glasses Company Blac... £5.60 £8/100g Amazon UK	The Reading Glasses Company Unis... £6.91 Amazon UK	Men's Women's Original Retro +0.50 +0.75 +1... £2.99 Amazon UK	Wraparound Safety Glasses £5.74 Nisbets UK	Folding Reading Glasses Compact Telescopic Silv... £12.95 Amazon UK	Safety Glasses Economy For Eye Protection Clea... £1.02 Zafaty Supplies

Organic results



See University of Westminster results for eye glasses >

Glasses Direct™ - 2 Pairs From £15 - As Seen on TV
<https://www.glassesdirect.co.uk>

Web Glasses Direct™ - 2 Pairs From £15 - As Seen on TV Free home trial Free returns Glasses Direct Help Login Excellent 196,701 reviews on Glasses Sunglasses Lenses Brands Home Trial Offers Help & FAQ 50% OFF ...

Men's Glasses
Designer Boutique at Glasses Direct
Delivery

Women's Glasses
A-Z Brands
2 for 1 Designers

EXPLORE FURTHER

Prescription Glasses from £6 with Free Lenses - SelectSpecs	selectspecs.com
Cheap Glasses: where to buy online - MoneySavingExpert	moneysavingexpert.com
2 for 1 glasses from £69 Specsavers UK	specsavers.co.uk
Asda Opticians Buy Contact Lenses Online Affordable ...	opticians.asda.com

Example. Google.com

Paid results



The screenshot shows a Google search for "car insurance". The search bar at the top contains "car insurance" and the Google logo. Below the search bar, there are tabs for "All", "Images", "News", "Shopping", "Videos", and "More". The search results are displayed below the tabs. The first result is a paid advertisement from MoneySuperMarket, titled "Compare Cheap Car Insurance - Compare at MoneySuperMarket". It includes a star rating and a brief description. The second result is another paid advertisement from Confused.com, titled "Cheap Car Insurance Quotes - 96% of customers recommend us". The third result is a paid advertisement from GoCompare, titled "GoCompare™ Car Insurance - Compare Car Insurance Today". The fourth result is a paid advertisement from CompareTheMarket, titled "Compare UK Car Insurance - Compare The Market™ Car". The fifth result is an organic search result from GoCompare, titled "Compare Cheap Car Insurance Quotes | GoCompare". It includes a brief description and a link to the website.

Google car insurance

About 1,770,000,000 results (0.38 seconds)

Results for Bournemouth · Choose area

Ad · <https://www.moneysupermarket.com/> ·
Compare Cheap Car Insurance - Compare at MoneySuperMarket
Compare **Car Insurance** Quotes at MoneySuperMarket and You Could Save up to £330 With Us!
★★★★★ Rating for moneysupermarket.com: 4.8 - 10,355 reviews

Compare Car Insurance
Various Types of Car Insurance. Compare From Different Providers.

Super Save Price Promise
We'll Price Match The Same Deal & Give A £20 Gift Card. T&Cs Apply

Ad · <https://www.confused.com/> ·
Cheap Car Insurance Quotes - 96% of customers recommend us
Get a **car insurance** quote today and see how much you could save with Confused.com

Ad · <https://www.gocompare.com/> ·
GoCompare™ Car Insurance - Compare Car Insurance Today
We Search More **Car Insurance** Providers Than Any Other Major Comparison Website. Save...

Ad · <https://www.comparethemarket.com/> ·
Compare UK Car Insurance - Compare The Market™ Car
Fuel A Better Deal On Your **Car Insurance**, Compare 56 Providers.

<https://www.gocompare.com> · car insurance ·
Compare Cheap Car Insurance Quotes | GoCompare
Compare **car insurance** quotes. It's quick and easy to compare **car insurance** and find cheaper cover – we just need a few details about you and your vehicle.

Organic results



User Behaviour

Eye-tracking results from the book 'The Art of SEO'

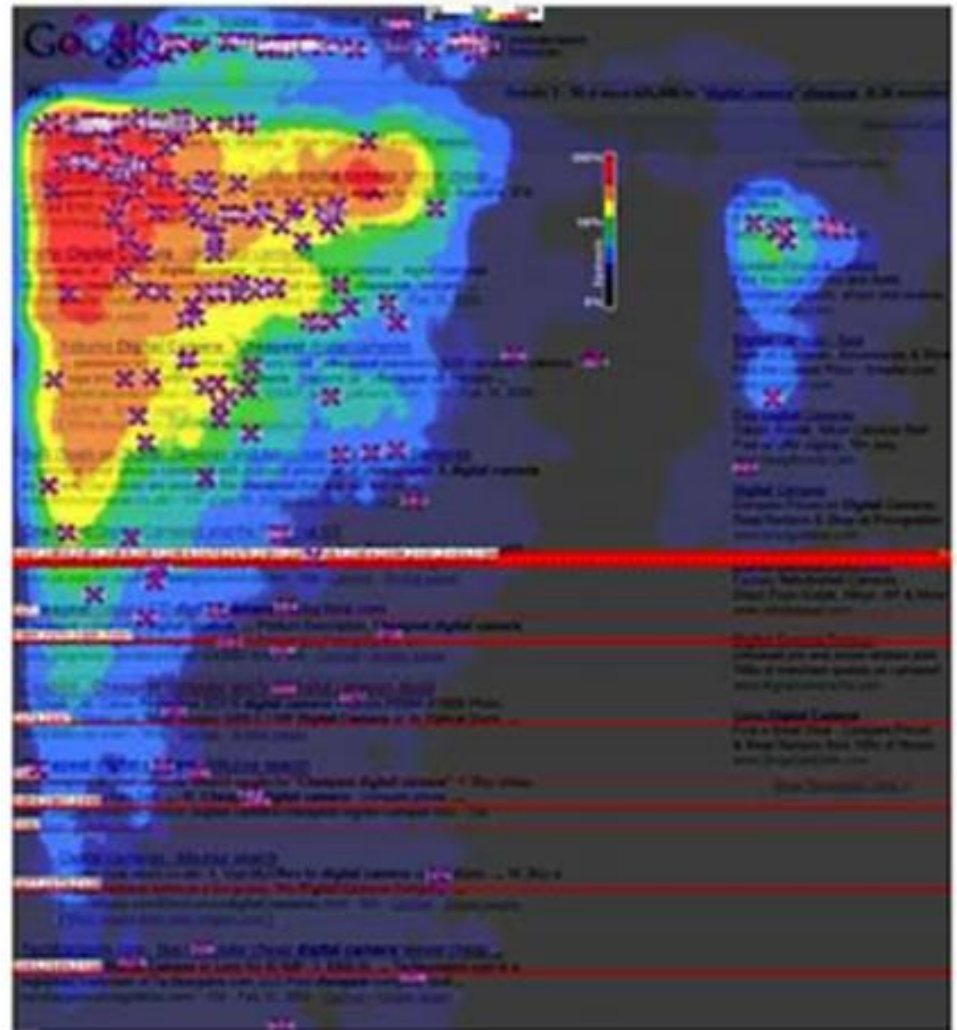
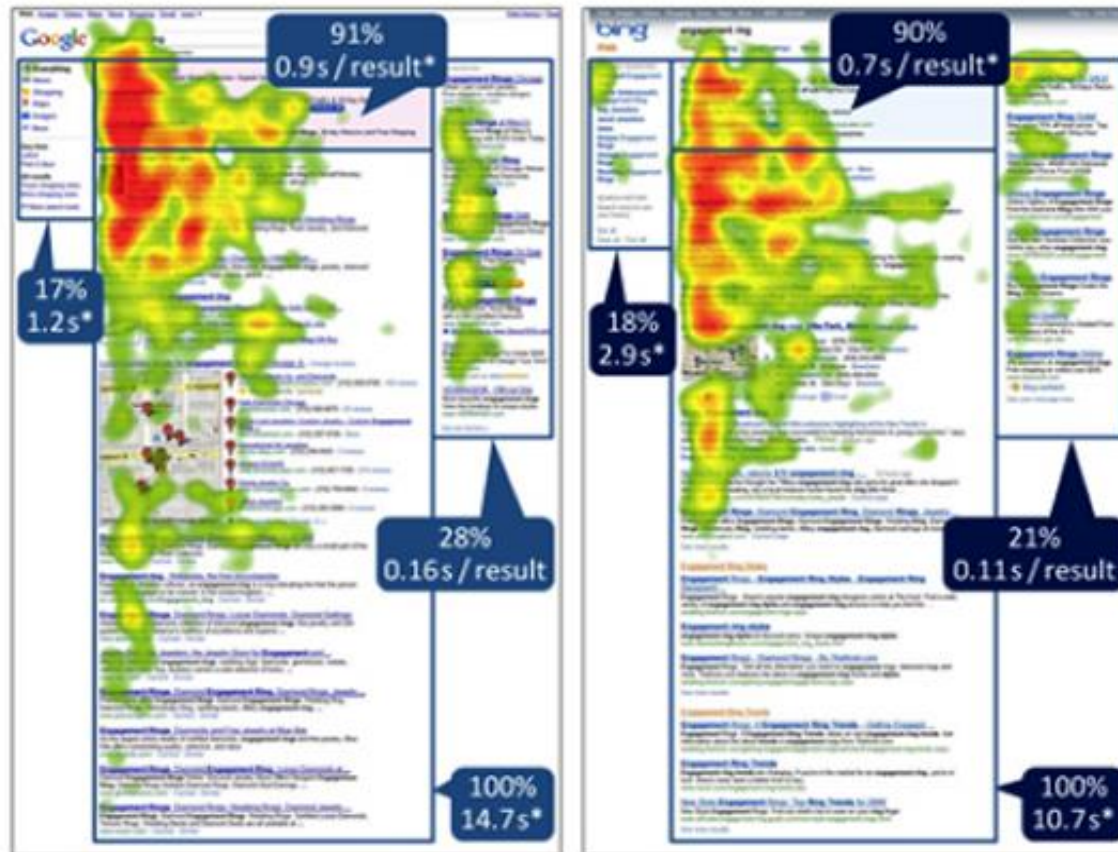


FIGURE 1-8. Enquiro eye-tracking results

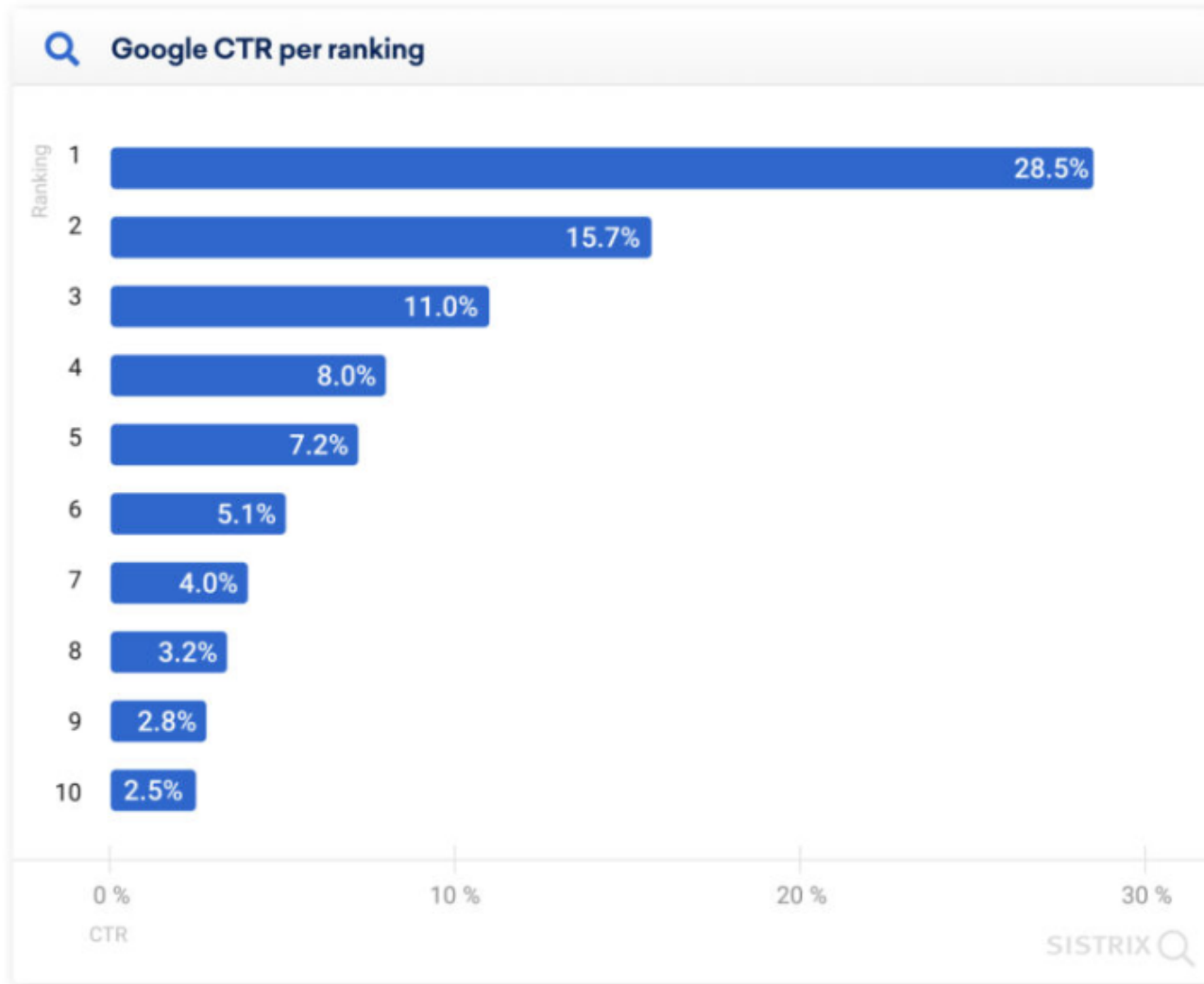
Heatmap of gaze time (Art of SEO)



Heatmaps showing the aggregate gaze time of all 24 participants on Google (left) and Bing (right) for one of the transactional tasks. The red color indicates areas that received the most total gaze time (4.5 seconds and above). Each callout includes the percentage of participants who looked at the area and the time (in seconds) they spent looking there. The numerical data are an average across all four tasks. Asterisks indicate values that were significantly different between Google and Bing at $\alpha = .1$.

FIGURE 1-10. User Centric eye-tracking results

Google click through rate



(Search Engine Journal, 2020)

Early SEO methods

- **2000-2010:** Most common SEO techniques
 - Keywords
 - Keyword frequency (number of appearances in a page)
 - Keyword density (appearances as a ratio of all words)
 - Backlinks
 - Meta – tags (descriptors added at the HEAD of HTML pages)
 - Doorway pages
 - Hidden text
 - Alternative text
 - Domain name optimisation (later in the period)

2011 till today

- **2011:** Panda (3 updates February, April and September – focus on removal of thin content)
- **2012:** Penguin (April 2013 – filter out sites that abuse SEO methods, e.g., spam links. Sites were degraded / devaluated)
- **2013:** Penguin 2.0 (May 2013 – authorship, social media)
- **2013-2015:** Hummingbird (Sept. 2013...
Rewriting of the main search code, consider context and meaning rather than just keywords)

A History of Major Google Algorithm Updates

<https://moz.com/google-algorithm-change>



Did 'primitive' techniques disappear?

- No...the question is **how** we implement those techniques.
- In the past the main question was about 'quantity' (e.g. quantity of links; quantity of content etc) – now it is more about **quality**.
- Developers also have to be aware of the penalties for abuse of the system.
- We have already seen how important SE are for acquisition of prospects.

The challenge for the SE...

WEB DEVELOPER: “The Crown hotel in London”

USER: “places to stay near Covent Garden”

[1] The SE links these concepts together through language (RELEVANCE):

Make a semantic and thematic connection between the keywords a user enters and the content of the website.

[2] The SE determines relative usefulness (RANKING):

The SE evaluates numerous factors, including:

- ☐ How the site’s content is written (its quality)
- ☐ How the content is organised and implemented
- ☐ What other websites are linked to it (backlinks)
- ☐ How ‘local’ a site is to a given user (geolocation)

Categorisation of SEO techniques...

We can identify **three** major areas / techniques related to SEO:

1. Keywords and site content [on-site SEO]
2. Site organisation and technical SEO (domain name; URIs, root domains & sub domains; structure; navigation; internal linking; meta tags – snippets; sitemaps.xml and robots.txt)
[on-site SEO]
3. Backlinks (external links) [off-site SEO]

NB: Optimisation via social media, would generally be considered under the 'backlinks' technique

1) Keywords and content

- What is a keyword?
 - Several definitions – depends on the perspective
- ❖ Keyword (from user's point of view): It is any word or short phrase that will be used to search for a topic (what the user types in search box).
- ❖ Keyword (developer's point of view): It is any word or short phrase they will use to optimise the site; a word or a phrase you would like to appear for, in search results
- ❖ Keyword (general definition): It is any word or short phrase that describes a website topic or a page.

1) Keywords and content

The main question for the developer/SEO expert is to identify the best keywords that will make the site visible to users *who are looking for this site* and then build the content of the site around those keywords.

Keyword research is the process used to identify keywords or key phrases relevant to a website or web page.

As part of this process, we try to understand:

- What terms people type in search engines
- How frequently they do it (volume)
- How relevant to the sites business objectives these terms are
- How competitive those terms are with other websites

NOTE: each page could have different sets of keywords

1) Keywords and content

Keyword research can be divided into 5 key phases...

- Phase 1:
 - Make sure you understand the different types of keywords and their structure.
 - Not all keywords will be relevant to the business's objectives and goals.
- Phase 2:
 - Brainstorm keyword ideas.
- Phase 3:
 - Evaluate your keywords by finding 'search volume' and 'competitiveness'
- Phase 4:
 - Categorise and filter out candidate keywords.
- Phase 5:
 - Develop your content around your keywords – develop your own style

Understanding keywords

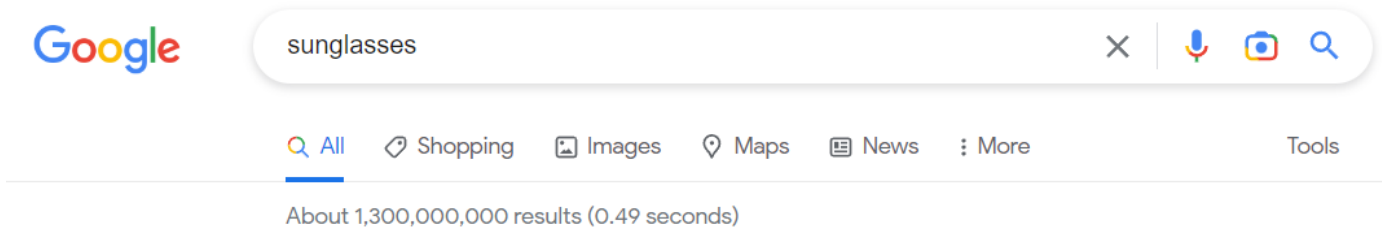
Lets considering an example...

- Imagine that you recently opened an online business selling high end designer sunglasses.
- The keyword “sunglasses” is a term that is searched quite frequently – someone might think this is a good term to use.
- On the other hand we should consider **the relevance** of the keyword to our website and our business strategy.
- If we only sell non-prescription sunglasses, then people who are looking to order lenses with their glasses prescription will be disappointed.
- Similarly, we don’t sell every brand of designer sunglasses. If the keyword is “sunglasses” we could potentially attract users who are looking for brands that we don’t currently offer.

Understanding keywords

Furthermore, there are several other websites (including large established retailers like 'Amazon' or specialist opticians 'Specsavers' etc) that will try to optimise with (rank for) the keyword 'sunglasses'.

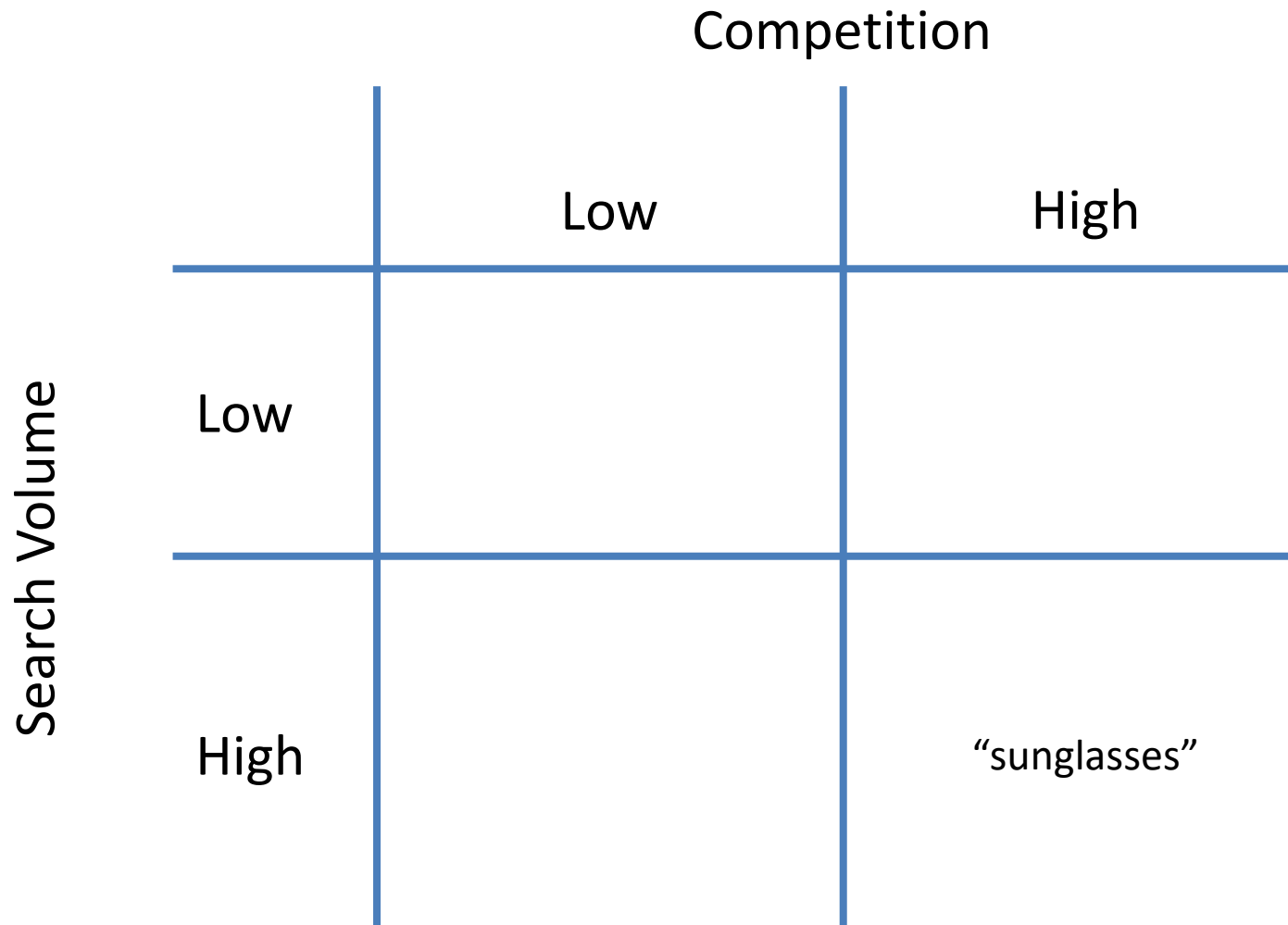
As a result, 'sunglasses' is a very *competitive* keyword.



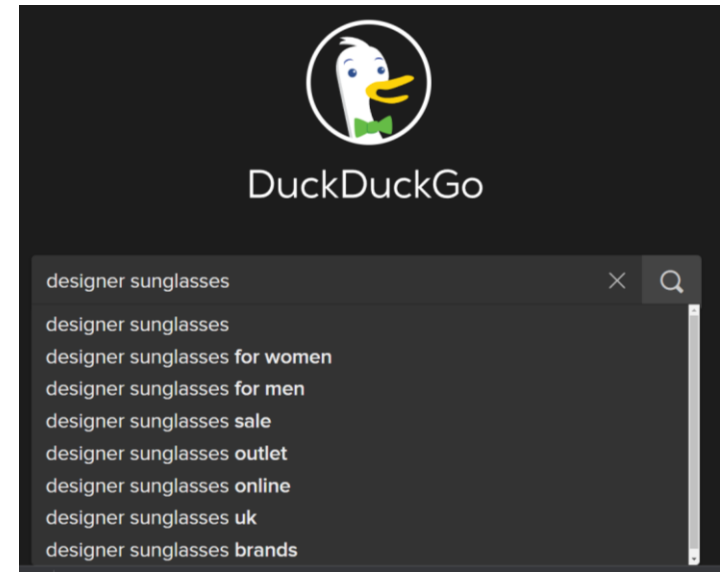
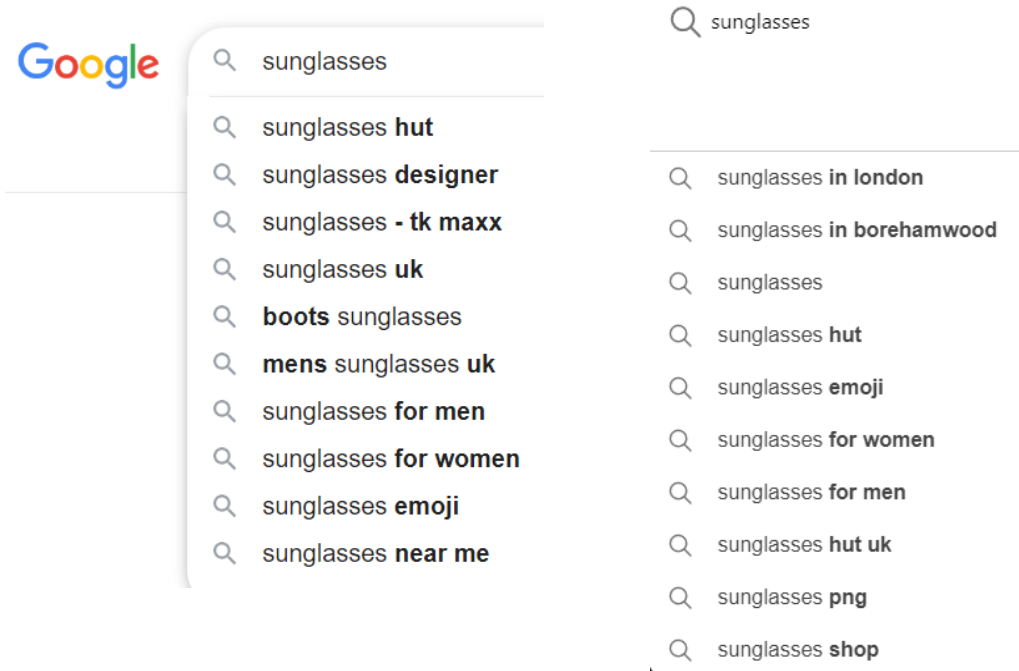
Ideally, we should target a keyword/key-phrase, that is

- ☐ Relevant
- ☐ Frequently used
- ☐ NOT competitive

The four quadrant system



Keyword ideas



One of the best ways to generate keyword ideas is to use the search engines themselves.

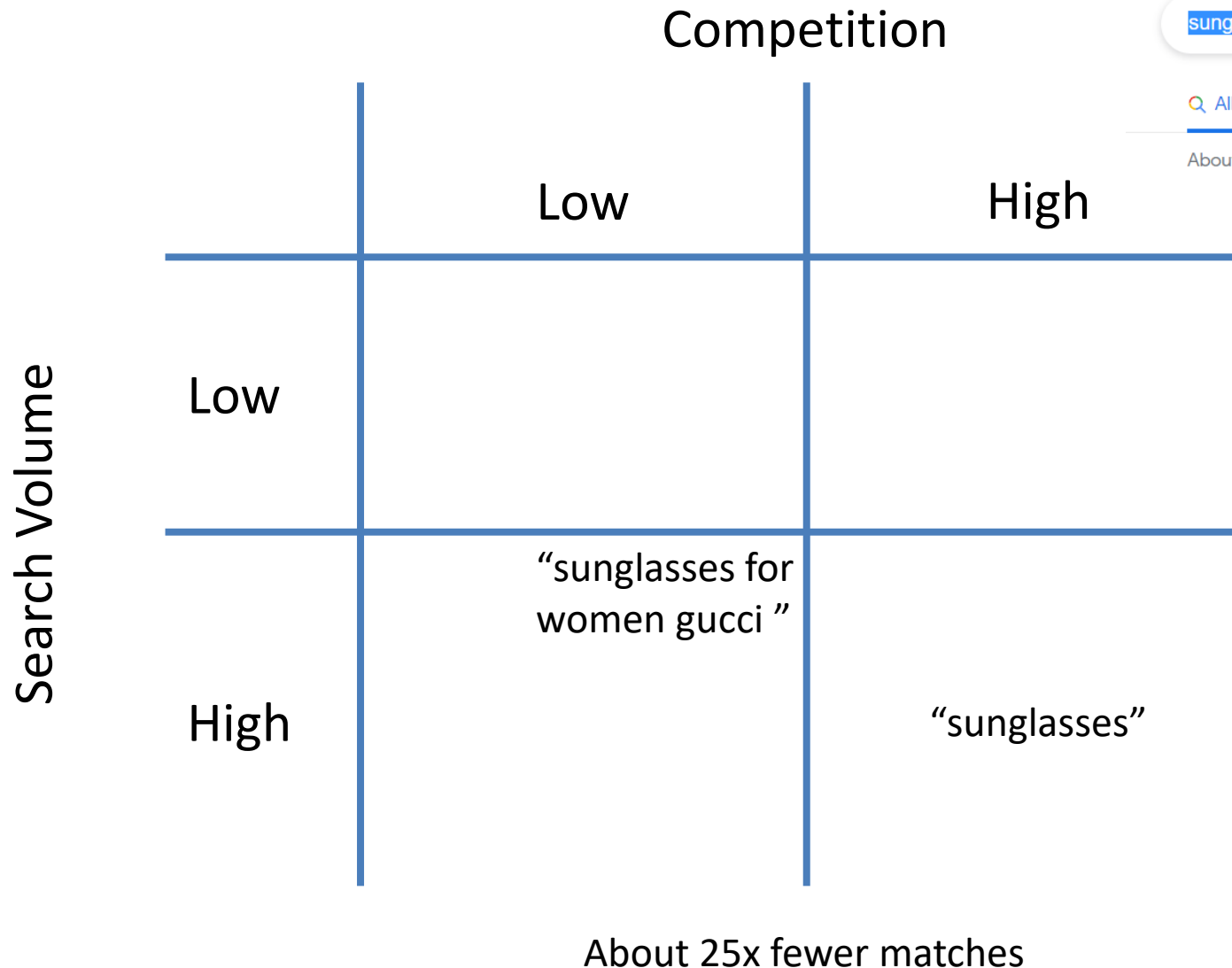
Understanding keywords

Apparently as you start typing your query in a search engine, the engine tries to make the keyword/phrase/query more *specific*.

Keyword elements (structure): *head / modifier / tail*

- ❑ *Head*: The focal point of the user's query and what the rest of the keyword relates to (e.g., sunglasses)
- ❑ *Modifier*: a single word or phrase that can change the type of search (e.g., sunglasses for women)
- ❑ *Tail*: a word that clarifies or adds further detail to the head term (e.g., sunglasses for women gucci)

The four quadrant system

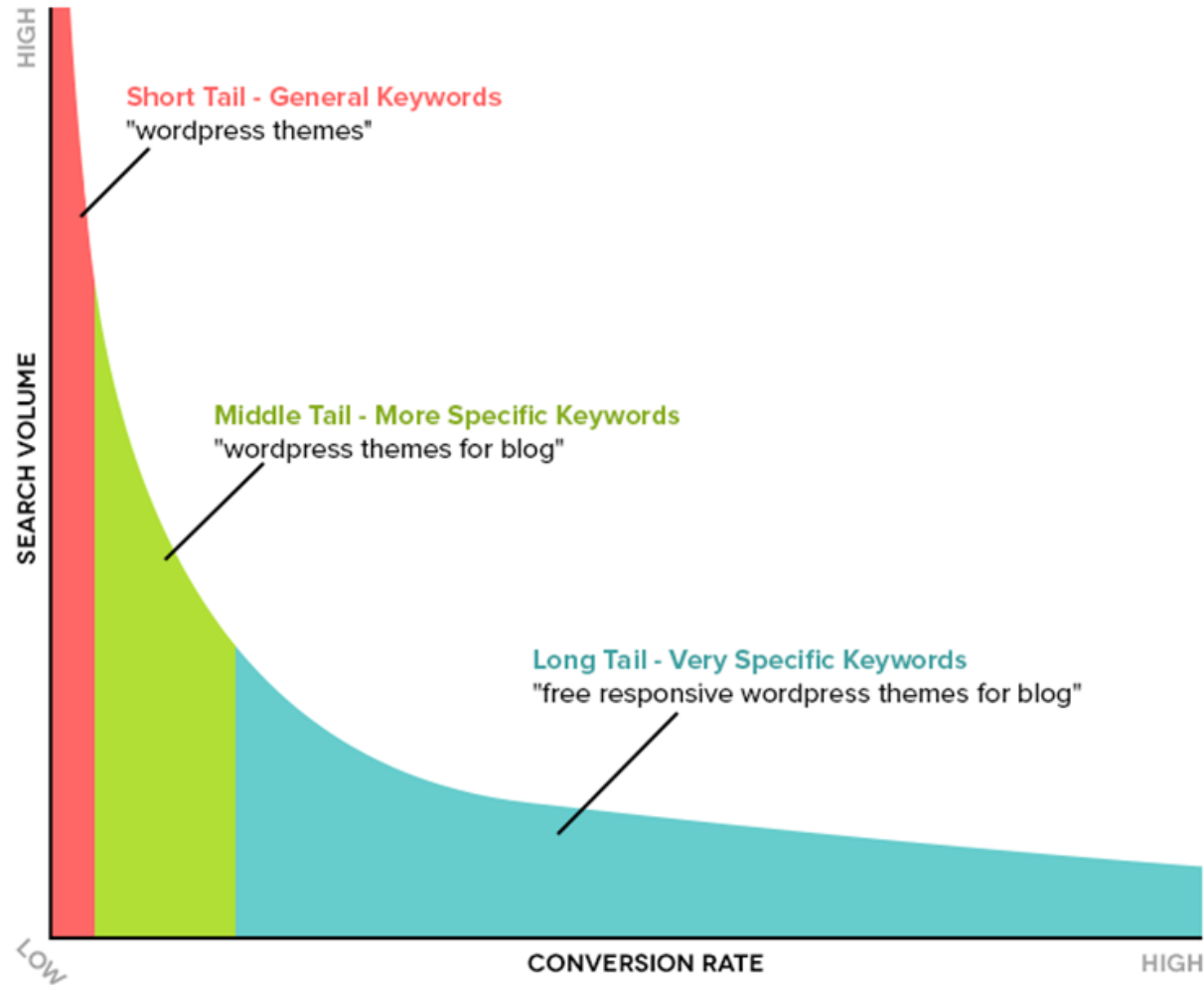


sunglasses for women gucci

[All](#) [Shopping](#) [Images](#) [Maps](#)

About 51,400,000 results (0.57 seconds)

Classifying keywords by conversion rate



Source: SEO Pressor

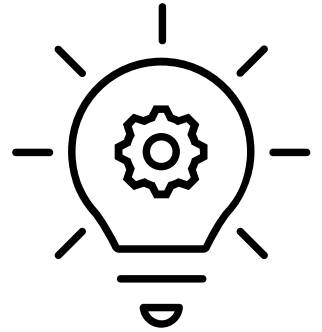
2) Brainstorm keywords

Create a list of potential keywords according to the products and services a website offers.

- Put yourself in the shoes of users.
- If a user searched the keyword and got to the site, would they be able to find what they are looking for?
- Would they be satisfied with what they found?
- Think about different ways the user might express the same intention
 - Vacations -> holidays
 - Low-cost holidays -> cheap holidays
 - Eye strain -> eye test
- We can also run competitors' websites through keyword density analysers
E.g., <https://www.fashionejewear.com/en-uk/collections/designer-sunglasses>
through <https://smallseotools.com/keyword-density-checker/>

2) Brainstorm keywords

designer sunglasses for men
designer sunglasses for women
designer sunglasses frames
designer sunglasses case
designer sunglasses case hard
designer sunglasses case soft
designer sunglasses with logo on lens
cheapest dior sunglasses
dior sunglasses women
dior sunglasses men
white designer sunglasses prada
do prada sunglasses have uv protection
are prada sunglasses polarized



3) Evaluate

From the list of **brainstormed** keywords, we check...

- ✓ Keywords lead to the 'correct' / expected results
- ✓ How frequently they are used in queries
- ✓ How competitive they are

Correct results: Try using them in search engines. What kind of results do they produce? Do you get back quality sites?

Volume: determine the number of monthly queries across the major search platforms (e.g., Google, Yahoo and Bing).

Competition: determine how many results the search engines provide for each of the keywords. Try to balance the **volume** with **competitiveness** according to the 4-quadrant system.

3) Evaluate

In this week's tutorial pack, there are some suggested sites you can use to determine search volume and competitiveness.

- ❑ <https://ads.google.com/home/tools/keyword-planner/>
- ❑ <https://trends.google.co.uk/>
- ❑ <https://www.wordtracker.com/>
- ❑ <https://tools.wordstream.com/>



4) Categorise

‘Categorising keywords’ is the process of grouping keywords into topics or themes.

By categorising keywords, you add meaning and organisation to the process of keyword research. It can also lead to something actionable, such as topics for individual pages or to understand which links to place on the homepage.

Following our previous example: we might place “*do prada sunglasses have uv protection*”, “*are prada sunglasses polarized*” in a common category around eye-protection.

5) Develop content

- With a set of relevant categories defined, it remains to develop quality content that makes use of your selected keywords.
- Quality implies original, informative, interesting and engaging.

Let's see an example. It comes from an article published on the internet. Can you guess what the author is trying to optimise the page for?

“Understanding Pomegranate Juice Benefits

Some people may not be that knowledgeable about pomegranate juice benefits but it is actually a very effective source of Vitamin C. The pomegranate fruit contains a lot of healthy nutrients and you can get a lot of good immune system boosters out of pomegranate juice benefits. It can actually provide around 16% of the required amount of Vitamin C that adults need to take on a daily basis. Pomegranate juice benefits also include Vitamin B5 as well as the antioxidant element of polyphenols and potassium.”

Site organisation and technical SEO

- This is a class of **on-site** SEO techniques whose purpose is to help search engine spiders to understand what a site is about.
- In addition, some of these techniques also help contribute towards ease of use and greater accessibility.
- The **main techniques** you should be aware of include:
 - Meta tags and code snippets
 - URI formatting
 - Alternative text and descriptions
 - Sitemaps and robots files
 - ...but there are many others

Meta tags and code snippets

What is a 'meta tag'?

A bit of HTML code that adds additional metadata about a site.
See https://www.w3schools.com/tags/tag_meta.asp

E.g.

```
<title> NFL Jerseys </title>
```

```
<meta description='Buy NFL Jerseys and NFL Jerseys and  
Football Shirts at the Official NFL Europe, Middle East & Africa  
Online Store. PayPal, American Express available.'/>
```

NFL Jerseys | NFL Europe, Middle East & Africa Online Store

europe.nflshop.com/stores/nfl/en/c/jerseys ▼

Buy **NFL Jerseys** and **NFL Jerseys** and Football Shirts at the Official NFL Europe, Middle East & Africa Online Store. PayPal, American Express available.

URI Formatting

This is the process of making URIs used on the site more search engine friendly (**less cryptic**) whilst providing a greater idea of what the page is about.

For example, which of the two URIs do you think is better?

☐ <https://www.bbc.com/article/192182102>

☐ <https://www.bbc.com/news/business/bank-of-England-to-announce-digital-pound>

Alternative text and descriptions

In HTML an image can be added to a page using the `` tags

- But what if an image fails to load?
- How does a search engine know what the image shows?
- What about people on low-bandwidth connections?
- What if the person is partially sighted or blind?

In this case we can provide alternative text that will be displayed in its place. This can also be read aloud by various **text-to-speech programs**.

E.g. ``

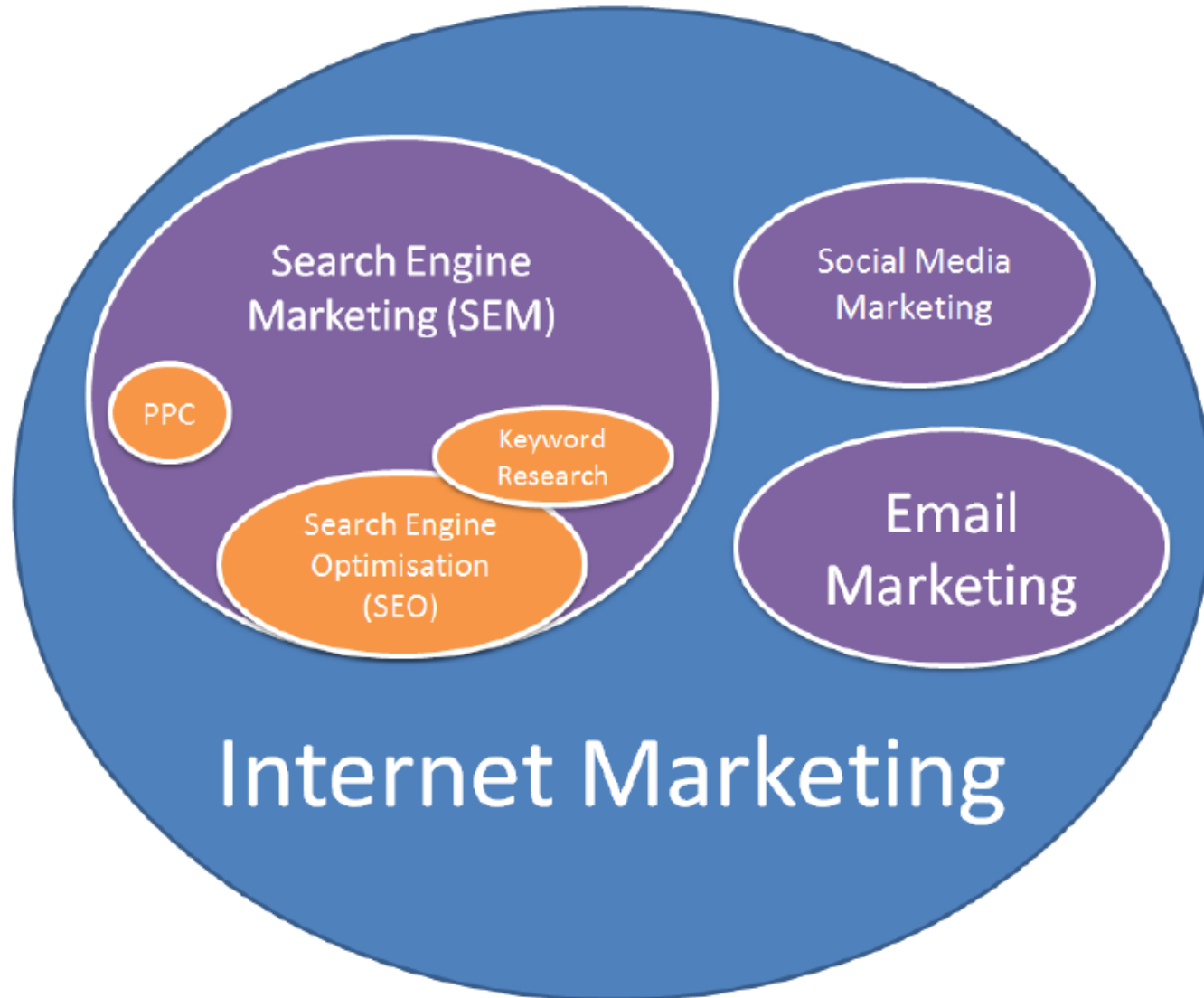
Sitemaps and robots files

A ***sitemap.xml*** is a file where you provide information about the pages, videos, and other files on a site, and the relationships between them. Search engines like Google read this file to crawl a site more efficiently. A sitemap tells SEs which pages and files you think are important in your site.

A ***robots.txt*** file is used to control how search engines crawl and index resources from a web site. For example, we can tell search engines to ignore certain file types or pages so that they are not present in search results.

Let's have a look at these files for the University of Westminster website.

SEO with the context of internet marketing



Pay-per-click (PPC)

In a nutshell

PPC concerns paying for traffic based on the number of clicks on a link to a web page.

In a nutshell

One of the most famous schemes is Google Adwords, but others exist including for Bing and Yahoo.

In a nutshell

Advertisers usually agree an amount per click or enter into a bidding process. In general, the one who is willing to pay the highest will be given the first opportunity for their link to be displayed.

PPC Advantages and Disadvantages...

- PPC has a number of advantages to traditional forms of internet marketing and SEO.
- Results are almost immediate and there is lots of options in terms of when links are displayed and to which groups of users.
- They can be highly contextual
- Weaknesses lie in the area of cost, especially for some very competitive keywords.

PPC Mechanisms

- The marketplace for Google Adwords (PPC) resembles an auction
- Why do you think auctions are used? Why not just specify a price?
- Advertisers bid on specific (groups of) keywords
- English auction? Dutch Auction? Sealed Price? Sealed-second Price (Vickery)?

Google AdWords

- In fact Google Adwords resembles a sealed-second price (Vickery) auction
- The dominant strategy is for each advertiser to submit a price equal to its value to them.
- In this case the advertisers have no control over how much they pay for the keyword, since it depends on the next highest bid, they can only influence whether they win or loose.
- Bidding higher increases chance to win but with zero or negative expected payoff
- Bidding lower decreases chance to win and leads to zero expected payoff

Determining the value of a click...

- Deciding the amount to bid is often influenced by the expected value of the associated click-through
- Suppose a charity website decides to use PPC to attract more visitors to its donation page
- They have calculated that in the last month there have been 38,000 hits to the donate.html page, 3,400 donations have been made totaling £45,500.
- How much is a PPC placement worth?
- First we calculate average donation £13.38 value and the average conversion rate 0.089%
- Multiplying these numbers we estimate the expected page value £1.19

In Summary

- Search engine marketing is a category of internet marketing designed to increase a websites position in the search results.
- We have already seen how search engines play a major role in the marketing funnel as they are a key acquisition channel, enabling the conversion of suspects into prospects.
- Search engine marketing includes strategies such as search engine optimisation (designed to influence organic results) and pay-per-click (which influences the positioning in paid search).
- A keyword is any word or short phrase that can be used by the search engine to match user queries with relevant sites.
- Keyword research is a process that concerns identifying relevant keywords that users might type into a search engine to find a site.
- It is important to strike a balance between search volume and competitiveness when selecting between competing sets of keywords

End