The Nuts and Bolts of RealTime Image Search

(www.nachofoto.com)

A Nachofoto White Paper

Published March 2010

For Additional information contact:
Vineet Agarwal - vineet@nachofoto.com
Anuj Agarwal - anuj@nachofoto.com

Users expectation from image search engine

When Tiger Woods made his first public appearance in February 2010 at a press conference to make public apology after 3 months when he crashed his SUV into a tree outside his Florida home, hundreds of millions of users across the Internet who typed in "tiger woods" in image search engines expected to see latest images of the icon. All of the traditional image search engines returned images of tiger woods published several years ago.

In another incident back in October 2009,when parents of six-year-old boy, Falcon Heene claimed that their son floated away in a homemade helium balloon flying at altitudes over 7,000 feet, millions of people around the world searched for this term expecting images of either the homemade helium balloon flying at 7,000 feets or of the six-year-old boy,falcon heene or of his parents.All the traditional image search engines returned totally unexpected images in which none of their million users were interested in.

In general, users prefer to see latest and high quality images for any given search term where recency is applicable.

Classification of Image Search queries

There are 4 different type of search queries that predominates image search engines.

Type1 query - "static keywords" whose meaning and their images do not undergo any significant change with time.New images published on the web for these queries are same as the old ones.
Examples

- balloon
- banana
- Niagara Falls
- Eiffel tower
- broccoli
- Plants,Animals,Vegetables

Type2 query - "dynamic keywords" whose meaning does not change but their images undergo significant change with time.New images are being published on the web on a regular basis for these queries.
Examples

- **Lindsey Vonn** - When she won the Gold medal at the olympics,this term marked its position in google insights for image search in "rising searches" column for 30 days.Most of the search engines returned old images of the alpine Ski Racer.Users expected images of Lindsey vonn's vancouver olympics images.
- **Olympics** - At the time of winter olympics in Vancouver, users using the term 'olympics' in image search would want to see latest images from this event and not the olympic logo.All traditional image search engines returned past and current years olympic logo.
- **iPhone** - When users use this term, they are looking for iPhone 3G or 4G pictures and definitely not the age-old first generation model.
- Britney Spears
- Barack Obama
- Tiger Woods
- blackberry smart-phones
- ford
- electric cars

Type3 query - "dynamic keywords" whose meaning and their images both undergo significant change with time
Examples

- **Droid** - Refers to fictional robots or star wars characters.When Motorola launched "Droid smartphone", users using the term droid on image search expected to see the new smartphone pictures.
- **Saints** - Primarily refers to a religious person.When American Football team "New Orleans Saints" won the Super Bowl, this search term hit google insights and underwent a "breakout".
- **Balloon Boy** - At the time of "Balloon boy hoax" incident, people expected images of the homemade helium balloon and of the six-year-old falcon heene and not that of kids playing around with balloons.
- **9/11** - When the NYPD released new aerial photos of the 9/11 attack in February 2010,people using this search term in image search expected to see new photos of the 9/11 attack captured from the helicopter by the NYPD and not the old ones which were returned by all the traditional image search engines.

Type4 query - "New keywords" which were never used before on the internet.
Examples

- Amazon Kindle
- iPad
- Lamborghini Murcielago LP640 or any new product or a automobile or anything which has a unique name/brand on the web..

Traditional Image search approach

Traditional Image search engine,including Google,Yahoo and Bing, etc., use little, if any,image information to rank the images.They rely heavily on image file name,alt/title attribute of the <img> tag and to a certain extent "text" very close to the image.

Their algorithm is crafted in a manner such that they are able to return high quality image results only for type1 and type4 search queries. Traditional algorithm protocols are not applicable if they were to return the same High quality image results for type2 and type3 search queries.

Nachofoto,a realtime image search engine, is focused exclusively on solving that problem.

How NachoFoto works

4 main factors influence our search results

1. **Recency Factor(freshness**) - Our semantic text-analysis algorithm help us determine freshness of an image on a given webpage.
2. **Image density of a webpage** - Number of High quality images interlinked with each other on a single or multiple web pages determines image density of a page.Pages with high image density are given higher priority in our search results.
3. **Inward links** - Higher priority is given to pages with strong internal inward links.Anchor text of the inward link is considered very important.
4. **Domain Authority** - Domains with fresh,family friendly, high density images are given higher authority.Such sites are given higher preference in our search results.

Factors like user friendly web URLs,image file names,image tag alt and title attribute,caption surrounding image,informative text on the webpage,use of multimedia related keywords in important tags of the page and many other factors influence our search results.

Comparison: Nachofoto with Traditional Image search Engines

Traditional search engines are very good at finding image results for search terms of type1 and type4.Quite often they return irrelevant(not fulfilling users expectations) results for type2 and type3 search terms.

On the other side, Nachofoto has been successful in returning high quality results for type2, type3 and type4 search terms.

Our Research team conducted an in-depth analysis of image search results for most popular search queries listed on "Google Insights for Image Search", Google Trends(hot topics and hot searches) , "Yahoo Buzz" and "AOL Hot Searches".
(**Note:** Only those queries were considered which are applicable to image domain. Keywords like "masters live coverage" or "the masters tv schedule" fall under video and text domain search queries respectively and were not considered in their analysis)
For most of the popular queries, Nachofoto search results were superior than Google,Yahoo and Bing image search results. Majority of these popular keywords were of type2, type3 and type4 query.

These three types of keywords constitute at-least 10% of the queries made across all image search engines. With web-image search traffic slightly over one-third the traffic of web-text search, market share for the above three type of queries constitute over 3% of the search market.
With 1% of the search market worth $1bn per year, 3% is a huge target to go after.