Q&A WIM 2023-2024

Gabriel Rovesti

Disclaimer

Each question is answered as precise as possible, looking at all questions present inside MEGA.

For each, the format is this:

- features
- usage
- pros
- cons

Considering we want to add value to the overall content, here will be also present the usability analysis of all the websites present inside MEGA. Because we like to make things *simple (for real)*.

Feel free to reach me to feedback over this file content. Also to thank, which does not kill me that much.

Summary

Written by Gabriel R.

Starter kit

Weaving

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, pieces of other web pages are taken and inserted inside a website, hence using their keywords inside of it, in random positions. This technique is particularly useful when inside a web page some rare terms are inserted, which means there will be only a few relevant pages over that specific topic or sets of keywords.

In general, this technique gives a global bonus to a website and is much appreciated from search engines. Also, this technique reduces the possibility of being penalized by TF-IDF, (which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency). This works to avoid keyword repetition, so they will be used more times without being penalized.

This measure considers only a selected set of words, called champions, to properly increase rankings over results inside search engines. This way, we use different champions words, quite rare even, hence definitely increasing the overall measure in a fairly automatic way to make interesting and attracting content for the users.

As a con, we are affecting the page content, so try not to put unrelated content to avoid users going away.

More critically, the technique could become dangerous if used unproperly, because this is basically black hat SEO, practices which try to trick engines' interest in a not-so-honest way. Potentially, even if rare, there could be the risk of duplicate content if done incorrectly, worsening the user experience if such content is assembled randomly from different sources.

Also, this can be considered as a form of keyword stuffing, meaning overuse of words attempting to manipulate rankings. Overtime, if using rare or unrelated words, the content can slowly become irrelevant, especially with algorithm changes or even keyword cannibalization (multiple pages of site competing against each other).

Dumping

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, we insert rare keywords inside our site, even if they are not related at all to our content: this helps in getting an high score, considering terms will be defined as rare as both frequency and importance (high TF-IDF, which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency)).

The technique can also be used to target long-tail keywords, for which are longer, more specific keyword phrases that are usually less competitive and easier to rank.

As a matter of fact, we are inserting terms not related to page content, which can make difficult keeping the users, even if it tries to facilitate access to the site. This happens because the page may be different to what they were looking for, losing trust from users.

Written by Gabriel R.

Repetition

This is a technique used in the context of getting more score for a web page when calculated by web engines. Just like the name says, it basically consists in repeating one or more keywords multiple times, with attention to TFIDF (which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency)), given spamming keywords is considered a bad practice and the site can be considered spam fairly easy.

In this way it generates the advantage of increasing the relevance of the page with respect to a single one or to a low number of keywords.

Consider it's easy to get penalized, given this technique is fairly easy to spot by said search engines crawlers, so it's important to pay attention to countermeasures, for example regularly update content to satisfy both users and search algorithms, while increasing TFIDF, diversifying keywords usage and possibly following SEO guidelines to avoid doing anything misleading/incorrect.

Stitching

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, it's simply copy&pasting intelligently from different web sources (other web pages, forums answers, etc.), then assembling the content to quickly obtain relevant content.

This technique is fairly useful to quickly gain popularity given it's fast to take and produce and gets chance of each page to get visibility via copied content and allows for more pages (so more diversification), another good thing which rewards the site from the search engines side. Of course, there needs to be attention on what content is copied, concerning content assembling and copyright violations without proper attribution, while being ethically wrong and again it's just a shady way to get more traffic.

Broadening

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, synonyms of keywords/complete sentences will be used, this way covering better the queries of specific topics and better satisfying similarity measures of search engines to give additional bonuses. A fairly easy example might be the following: if one looks for Disney inside the Web and inside our page there is Winnie the Pooh as keyword, the site will get a bonus thanks to this technique and Disney was never written once.

This is seen pretty positively by search engines, helps also users considering it offers them more precise information and gives extra bonuses when similar keywords are employed, increasing semantic analysis by engines. As a con, some keywords might be diluted (excessive use towards specific context) and not be so versatile if phrases are not posed precisely, causing ambiguity over time. This fact alone might penalize TFIDF, and one should consider that.

Spam techniques

Cloaking

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, there are two version of the same site, one for the search engine bot and one for the user. If it's the crawler, a version of page containing the keywords (the "empowered" one) is shown, otherwise to the user we show the normal page they want to see.

Pages allow to recognize bots or humans simply by using automatic means to do just that, for example mandatory recognition banner "Are you a bot" from the search engine.

This is a really powerful technique, considering it seems like a win-win: we are not touching users pages content and search engines give us more score and this is really hard to spot, considering probably only human control could possibly find it. The page is more visible and gets higher rank pretty easily via this technique.

As a con, the penalty is really high, considering search engines will ban us if this is discovered (not appearing inside search results anymore or even indexing), given this violates search engines guidelines and it's a deceptive practice overall. This practice is called black hat SEO, a set pf techniques which violate search engine guidelines and are intended to deceive search engines rather than provide genuine value to users.

Anchor Text Spam

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, keywords are inserted inside the anchor text, so they are part of the body page, but treated as separate. This way, search engines will value the words inserted in anchors differently, given they are formatted differently (underlined, distinct colors, etc.) giving these words an higher rate and score and bringing in organic traffic. This also gives special scores to the target pages, with less limitations with respect to TFIDF (which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency)).

Anyway, keywords inserted in anchors can penalize a page (penalty filters, for example consider search engines guidelines, with the aim to provide high-quality context and decrease spam), given anchors are supposed to describe and give an idea of the pages the point to. It also happens to touch the real content of the page, possibly in a way the user doesn't like. Overtime, this can bring only short-term gains in SEO and possibly not so versatile to search algorithms evolution.

URL Spam

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, keywords are inserted directly into the URL address of the web page, so search engines will give added bonuses to this practice, like might happen for example when inserting links inside the anchors (so-called anchor text spam technique); also, it can be used in combination with other spam techniques.

This is a fairly easy technique, since it does not affect pages content but has to be employed carefully, given otherwise it's considered spamming, specifically when inserting a word multiple times, hence being penalized in the SEO rankings from TFIDF (which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this

logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency)).

Body Spam

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, we simply are inserting words into the body of an HTML page, shown directly to the user and touching directly and simply the content of the page. While it's simple, it's definitely something that touches quite consistently the page of user, accepting compromises on th -e TFIDF side ((which basically measures how much important is a word in a page, considering how much the term appears (TF = Term Frequency) weighting this logarithmically to give a fair measure, measuring the inverse of the frequency (IDF = Inverse Document Frequency)).

Given this can be spot by search engines employed, to avoid changing user content, there is a series of techniques called "hiding", which basically hide the "trash" inserted with spamming. This is particularly useful in the context of these techniques, since one can easily use the style of a page to make content not appear inside pages, for example putting text spam the same color as the background, very small images (1x1 px) to trick invisible clicks or also employing redirection, which uses JavaScript code to address the user towards other pages, but has to be employed carefully since again, spiders may find it, even though with code usage it takes time to be analyzed.

This last one can be further improved by the usage of cloaking, where users see a "normal" page, while spiders see the "empowered" one, so the one with spam content. This is hard to spot, but also has the highest penalties overall, such as not being displayed anymore inside web results and against search engines guidelines; the practices seen above also break guidelines.

Title Spam

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, keywords are inserted simply into the title tag of the document, in a fairly simple and effective way not touching excessively the content of the page itself considering the user will never see such modifications: of course, one has to be careful, considering keywords should be put properly to avoid problems with this technique.

It's also seen positively by search engines, considering the title is a special tag and it's given more weight in SERP results and search engines spiders to not give particular attention to that. There is again the risk of keyword stuffing, so the title may appear unnatural and mismatch the content completely if not done precisely, given this is one of the first thing a user sees and makes him click.

Meta Tag Spam

This is a technique used in the context of getting more score for a web page when calculated by web engines. In this technique, keywords are inserted into the metadata of the page (so called meta tag). This is a fairly easy technique without altering the page content, but this technique is very much abused, so search engines won't give any more weight to words repeated even several times, considering the majority of sites overtime did this.

As other techniques marked, this can risk keyword stuffing, being penalized overall in rankings and may need to adapt overtime to algorithm changes.

Problems

Ad banners positioning problem

Considering advertisement is one of the most hated things by users, it's important to place them well, possibly in a seamless way, to avoid users going away from our site because of ads presence. Specifically, one can try to present them in an interesting way, around the content the page revolves around or just making one more attractive, so the user will avoid them. Subconsciously, the most important content inside a web page is text and so in the scanning phase the user conducts in a page, the ads will be "defensively" avoided in an "escaping" algorithm defined as "zapping effect", so the user will simply go away from the image and from the ad itself.

To avoid this effect, we try to confound the user into tricking him think the advertisement is also content and so will be motivated to not go away from it. A good example of this practice is definitely blending, so the ad will have no borders (border effect – separates ad from content clearly – not good) and becomes integral part of the site. To make this particularly effective, it's advised to use them in combination with text, so the user attention will be naturally devoted also to ads, but also avoid fancy effects to distract him (specifically something moving or leveraging other forms like audio; definitely to avoid popup ads, another very distracting and annoying thing – if present, at least tell the user easily how to close them).

Tricks in this context involve using naturally beautiful people, sexualizing the content just to make it more provocative or even the complete opposite: using "normal" people where paradoxically some "not-so-ordinary" content, so the user will be impressed in the good way. In any case, avoid using bright colors or something distant from the text, so this would separate ads from content even more. Another important aspect is definitely their size: bigger banners tend to be seen more easily than smaller ones, but avoid them covering context, otherwise we get the opposite effect.

Also, from a first glance, they should immediately communicate what they want to: avoid slow loading and don't make the user click fraudulently on them. They should have a positioning following the user attention map, so in descending order good positions are the left column, the top of the page and the right column.

Distract the user into tricking him it's useful content may revolve around understanding user habits; this goes on the way of behavioral advertising, so they are ads that try to offer relevant/helpful content to the user; if the user understand there's useful content inside ads, timers of users diminish up to 40% and return rate can get to 80% easily. Another simple trick is using mini-games, which gamify the user perception of ads and actually get a good effect both in timers and return rate.

Only to mention, SPARQL is the base for other modelling models for data, like Dublin Core, which describes basic properties of documents and FOAF, which describe base properties of people.

Lost in Navigation problem

The lost in navigation problem occurs when we want users to be aware of where they are within a specific site being conscious of their relative position: this can be problematic in the current place but also on where to go next (the so-called “where” axis); this can happen because the page has a layout which is not totally clear by the user, it doesn’t display the position (e.g. breadcrumb) or because it’s a page with a completely different layout. From a usability perspective, this is one which can be expressed as a persistent problem, given it’s one of the common problems present in navigation.

The user, many times, can feel overloaded and wants to move quickly; so, changing color to visited links is the simplest possible thing that one can do; consider also the most used navigation movements, which are clicking a link or pressing the back button. The user always has to know where it has been before (visited links), where he is now (breadcrumb as said, which can display the location, the attribute [logical categorization] or the path [required to come into the page reached]) and where to go next.

Semantic Web

LOD (Linked Open Data)

(Exact question: Describe the LOD classification: its features, uses and potential advantages and drawbacks)

The generalization of semantic web data is Linked Data, which are used to feed knowledge graphs; a special type of Linked Data is LOD – Linked Open Data, which are linked data available to anyone for free. LOD are classified in a ranking from 1 to 5 stars (to have at least 1 star, data must be on the Web – order of numbers is the star ranking, so point (1) equals (1 star)).

1) Data available on the web but with an open license (open data)
    - Example: Images
2) Data available for free but in a structured format (machine-readable)
    - Example: Excel
3) Data available for free using a non-proprietary data format
    - Example: CSV
4) Data available for free and are presented in Semantic Web Format and URI identifiers are employed so that it is possible to point to a single piece of data
    - Examples: RDF/OWL
5) Data available for free with data linked to other data sources to provide context
    - Example: Graph

Starting from 3 starts or below, one can reach 4 or 5 starts via lifting, which consists in giving data a semantic format and this is done via various tools, which basically join different knowledge graphs and calculate the links between concepts using similarity measurements. This can be considered a pro of this technology (just mentioning: the converse would be lowering, so decreasing the stars rating).

For example. we can quote Open Calais (now called Refinitiv), technology which takes a text or a page and automatically recognizes it, tags it and allows to take useful information (e.g. keywords, links, text, etc.) or even Wikido, which basically lifts semantic information from selected websites in the right way, providing a useful site to the user.

The main con in handling this data is how much of it is present: usually, we talk about big data analysis, which requires computational power and powerful tools, not always accessible to the user, because not everyone feels confident interacting with a DB via an interface or because of lack of knowledge.

A way to retrieve data consists in offering a SPARQL web service, with specified endpoints, which allows to specify default/named knowledge graphs which can be used in the query.
Other examples to quote are DBPedia, essentially the semantic version of Wikipedia, which uses proprietary and different ontologies (classes of information, giving meaning to objects and abstracting from the syntax to define semantic) or schema.org, defining the most important ontologies to query data and gathering it all, defining the overall information.

RDF/RDFS

RDF (Resource Description Framework) is a universal language to express semantic information describing relationships and concepts formally in a machine-readable way, allowing to structure information while removing ambiguities.

It's composed mainly of triples, mainly composed by sentences of "subject-predicate-objects", representing data as URIs or literals (strings). It's more of an "enriched entity-relationship" knowledge model, so concepts can be linked visually via so-called knowledge graphs, representing links. As a model, it can either be written as N-Triples (with the structure just described) or via XML, which elevates complexity because it allows, in its various dialects, to elevate the complexity of language creating harder sentences, while also enriching the meaning of sentences.

The first of two is chosen, because the existing dialects of XML make information aggregation impossible and, by itself, this technology ensures interoperability between metadata, relationships and objects, without the need of language translation: this can be seen as a pro of this technology. As said before, another advantage comes from aggregated information, which can be seen via knowledge graphs, and this allows to link multiple resources identified by the same URI.

RDF is not enough however, considering it gives only the basic layer, so the basic model and merging for aggregation. We also need classification of information (via the so-called ontologies, which basically classify labels/classes of objects proving a syntax/meaning to objects) and an integrity check of data, allowing deductions between them in a minimum computational cost.

This is allowed by RDFS (RDF Schema), which defines objects via classes, sub-classes and individuals and verbs via properties, domains and intervals (ranges), allowing taxonomy via the usage of ontologies. Given RDF allows for automatic aggregation via URIs, a concept can take multiple meanings (problem already foreseen at the dawn of URIs conception by Tim-Berners Lee, hence the birth of Web Axioms) easily and this can become problematic (URI Variant Problem: different URIs for the same concept and URI Variant Law: usefulness of URIs decreases with the number of variants).

There are extensions to RDFS coming in form of vocabularies to support for example ontologies (OWL – Web Ontology Language), mapping concepts and imposing control decidably over them, or dictionaries able to express data semantically in general (SPARQL), about people in social web (Friend Of A Friend/FOAF) or linked data, useful to build graphs in the Semantic Web, or also about web pages (Dublin Core/DC).

As pros, RDF ensures semantic expressiveness between concepts formally and easily reducing ambiguity, while achieving interoperability between different kinds of information without the need of language translation, and thanks to its decentralized nature and ease of distribution, anyone can create a vocabulary or publish data about other resources. Thanks to knowledge graphs, representation is easy and definitely modularly adaptable towards other usages.

As cons, RDF is very abstract and verbose, so it's difficult to operate with manually; it may also face still challenges in fully capturing the intended meaning of certain concepts, performing deductions with a computation cost which can cause overhead in certain situations. Also, RDFS requires some knowledge about the basic details of what are URIs/triples/knowledge graphs, etc.

Written by Gabriel R.

Dublin Core

Dublin Core is one of the dictionaries first used to try to structure the Web in a semantic way. It's basically the standard to define base properties about documents/URI pages, specifically defining data like subject, description, publisher, contributor, date, etc. up to 15 basic and optional informative elements. This allows to extend RDF (Resource Description Framework) representing semantically the pages in triples on which data can be easily linked together via knowledge graphs or which queries can be made. For example, using meta tags it is possible via the keyword 'DC ' make these triplets available directly to the web so that intelligent bots derive useful information.

An advantage comes from its standardized elements that facilitate the description and organization of web resources (specifically inside libraries and archives), good community support worldwide, enhancing interoperability towards different systems and applications with wide adoption and allows for easier semantic structuring of pages in an interconnected manner. Cons come from granularity of concepts: some concepts may be too specialized to explain via this technology and might not be adapt to technologies diffusion and expansions.

FOAF (Friend of a Friend)

FOAF, defined as Friend of a Friend, is defined as the standard ontology for the social web, describing the properties of a person, for example a class Person with Name, surname, workplace, phone, etc. It's an RDF vocabulary (this one is the Resource Description Framework), used to describe social networks and relationships for individuals/organization in a machine-readable way in triplets (composed of subject-predicate-object to describe concepts) for ontologies (so, to classify data syntactically, providing meaning to objects).

This is pretty much interoperable given its machine-readable format, expressing in a straightforward way relationships, affiliations, attributes and connections. It's also got a decentralized structure, allowing users to control and manage their data easily, integrating seamlessly with different web technologies and is extensible, given custom properties can be added according to the situation.

The main cons of that come from its limited adoption, reducing the effectiveness in interconnectivity between social graphs, but also the lack of standardization and security concerns when sharing information to machines.

OWL

OWL (Web Ontology Language) represent the extension of RDF Schema with a semantic layer for ontologies, which basically extend the semantic representation in N-triples of concepts given by RDF, classifying objects in classes and subclasses, while at the same time providing taxonomy and meaning via classes of objects (ontologies themselves). In the Semantic Web, a resource can possibly be present multiple times via the same URIs (problem of the URI Variant), so this was mainly added because of that.

OWL is infact a W3C language used to connect vocabularies using relationships between concepts, specifically indicating when classes and properties are equivalent or different, where functionally have the same features, or cardinality of relationships. In fact, as an advantage, consider the enriched schema to allow even more precision and categorization of concepts. Infact, it is designed to be highly expressive and represent information in a machine-readable format, while at the same time being formally correct to allows automated processing and inference.

Written by Gabriel R.

Its properties are expressiveness and logical decidability; OWL has a polynomial complexity, but it can also lead to more complex queries, which evidence the limits of OWL, better suited for static knowledge representation, also complex for not technical in this field. It's not possible to have high expressiveness and decidability at the same time and this guaranteed by OWL extensions. In order of increasing power, we have:

- OWL Lite, limited but decidable expressiveness
    a. complexity inside the so-called SHIF logic – this has $EXPTIME$ complexity
- OWL DL, less limited in expressiveness, always decidable
    a. complexity inside the so-called SHOIN logic – this has $NEXPTIME$ complexity
- OWL Full, which exploits more advanced logics, has no restrictions (so, implies high expressiveness), and it is undecidable

SPARQL

SPARQL (SPARQL Protocol And RDF Query Language) is basically a SQL-like for semantic data, so it allows to make queries inside of that. Computationally talking, it's decidable, so it terminates (polinomial space complexity- PSPACE) but is limited in power to a similar syntax to SQL.

It's structured in triples, made each of "subject-verb-object", making a pattern-matching between words and reasoning with knowledge graphs, used to describe information precisely. Some keywords used in the language include: PREFIX (used to create readable names from URI/URL namespace) and usual ones like SELECT, FROM, WHERE, ORDER BY, etc. allowing all operations of retrieval of data.

SPARQL also allows to deal with possibly partial information present inside the Semantic Web (given knowledge graphs usually combine information from multiple sources) using the keyword OPTIONAL, which means data can possibly exist or not and this way, the query won't give errors. The main disadvantage comes from particularly complex queries, which can possibly take much time to execute. This is mainly caused by OPTIONAL, because we may fall into co-NP class.

Link structures

PageRank Optimal Structures

(The exact question was: Describe at least 1 optimal structure of an alliance to optimize PageRank)

In order to boost the hypertextual score of a page people can use different techniques to create flows of data which will redirect to other pages, called target, from other ones, called empowering. This comes both from ingoing links and also outgoing ones, having the least possible number of links while keeping reachability from search engines spiders. If a spider reaches the target page via an empowering one (given there is a bidirectional links), all the pages get reachable so the search engines will see them.

To define correct spam farms, so structures created to increment the score and ranking of pages, we use various kinds of alliances and schemas, like:

- deep alliance, which given 2 target pages owned by different entities, double-linked with its pages and the allied ones, the average of the PageRank between them, implying robustness and stability

- superficial alliance, which given 2 target pages owned by different entities shares all unidirectional links of the 2 pages double-linking the 2 targets, creating a vortex of flow. The solution is more robust and maintainable, given there is only 1 double link between 2 targets, hence minimizing the number of links. Each target increments proportionally to the allied pages so each target will have a score greater than the maximum of original target score

- ring alliance, which considers 3 of more target pages owned by different entities and the target pages index in a circular path, so shares all unidirectional links of the creating a unidirectional vortex of flow between the targets. This way, the score of each target is greater than the original target score and it's powerful, given we link each page

- complete core, like the previous one, but creates a bidirectional vortex to basically anyone. This empowers ring spam farms considering a bidirectional structure and each target page as ingoing and outgoing links towards near page and, this way, the score of each target is greater than the original target score and it's powerful, given we link each page

Search engines try to avoid strongly connected graphs, which are structures like the latest two, given they are the strongest ones and often try to employ modification of their algorithms or control the shape of links to countermeasure unbalanced pages.

Link structure

(Exact question: Describe at least one Link Structure listing its characteristics, uses, advantages and drawbacks)

Good question – what exactly is the prof referring to? From the Telegram group I found:

"I think he wants to hear something regarding spam farms. For instance one of: deep alliance, artificial alliance, ring alliance, complete core alliance"

Personally, I think it's right, but one should also completely quote and talk about PageRank/inlinks/outlinks, then expanding of what was written above

(Onto the answer....)

Link structures are related to the hypertextual component of a page, related to the PageRank algorithm, which was one of the first algorithms Google revolved its working logic around. The principle is simple: the more links there are, the higher the rank of a page.

Basically, the sum of every pagerank has to be 1; subsequently, a randomization was introduced to give more balancing to the algorithm, introducing the concept of Markov chains, in which the so-called spiders browse the web and calculations are introduced to understand how much a specific link was clicked. There can be problems however; spider might get stuck inside pages (spider traps) or reach links where there are no external links (island).

Because of these, a teleportation factor was introduced to give more democracy in link selection for pages, solving the problem already seen; other ways to solve this problem include the usage of techniques like Totalrank where all pageranks are summed together as integral, giving an average to all links – again, this technique might be unbalanced if links are not cooperating.

For each page, there are measures of the types of links, specifically the ones going out (called outlinks) and the ones that take to our site (called inlinks). Over inlinks, to increment the pagerank of a page, we should have it aimed towards the highest number of external pages. So, the main con is how these inlinks are handled; possibly, they might introduce unbalance lowering the overall score of pages, otherwise they can be very effective, introducing techniques like:

- infiltration, so one "infiltrates" in various sites and tries to insert links to our site
- honey *pot*, creating "yummy" content (useful/good one), naturally receive incoming links
- link exchange, basically "joining forces" with other websites to exchange links
- resurrection, so buy defunct websites with a pagerank high enough.

There are also outlinks, which basically go outside our site and give pageranks to others; while this can seem not so effective as first, it guarantees solidity, given the spamdex is not only local. Thanks to the teleportation factor, pagerank can increase randomly with unforeseen consequences.

To avoid these, there exist different techniques to increase pagerank of both categories of links:

- spam farms, which are structures of links and pages consisting in pointed pages (target) and bidirectional links to reach them (empowering) trying to always keep reachability. They are considered optimal if search engines spiders reach a target page via en empowering one

To properly exploit this, there exist alliances of links, with different schemas to consider:

- deep alliances, where the target pages redistribute empowering pages flow of another site creating a stable pagerank (the average of two pageranks)
- superficial alliances, using a single point of contact between two pages and the pagerank is obtained as more than the max between the two pages links, giving a bonus in flow
- ring alliances, which as comprehend more than two websites and the target pages index in a circular path the next target page creating a sort of ring. If one of the internal pages crashes or has problem, this can cause problems to all other pages
- complete core alliances, which are bidirectional structures and each target page as ingoing and outgoing links towards near page

While alliances are generally a good thing to empower websites, to avoid unbalance, search engines try to fight complete core/ring alliances, which are the strongest ones, calculating the spam mass of individual pages (how may secondary pages are offering contributions to a main one – if it is too big, something is wrong) or analyzing the shape of websites, considering they are usually papillon like, so pretty balanced towards links redistribution – if it is too much different from average, again, something is wrong.

Websites Analysis

For each of these website, assume there is a static image of the site even if I put the link, just like you would find inside the exams. These ones are present in chronological order and here we try to do our best on every single one of these.

Note: These ones are personally made, not proof-corrected. In case, use them as logical reference. I don't assume them to be correct according to what the professor would want.

Santa Claus Village

(Website reference: https://santaclausvillage.info/it/) – Here 2023 version was analyzed

The website has already a clear enough name under its belt: it clearly communicates what it's about, but already does not utilize the .com domain, but instead a .info one, which can distrust some users.

For a broader view of information redistribution and analysis, we represent the following axes:

- the "who" axis, in which we can clearly understand what the site represents, specifically with the website name and logo on the top-left side of the main page. Also, social links are present
- the "what" axis, which the site displays quite well, with descriptive images about the Santa Claus context and different links about related activities (e.g. experiences/shopping/eat and drinking, etc.). There are multiple elements accommodating the finding of new information, possibly intriguing the user to look out for more. The user can also give info about opening times and even the map, to get new information each time. Also, the carousel of images seems related to the season content, possibly giving the user more information
- the "where" axis, which here seems to be somewhat respected, given there are different occasions and menus for the different seasons and also different options in menus to have both dropdown menus and links
- the "when" axis, where no link is visible to get the latest news, so we have no idea if the site was or has been updated in a while
- the "how" axis, respected by a search bar accommodating most of the internal search inside site. Main menu and content accommodate moving towards other things, which is not bad overall
- the "why" axis, giving the user so much content and choice to possibly get to the next link and information of choice and also a different range of services to consider and possibly redirect the user. That is the main reason to stay on site, possibly getting something from all of these services, which anyway are collectively presented

Overall, the site, as itself, presents quite nicely and conveys its information. As a side note, all the page content is taken by an image and not text, but the informative text over the banner is positioned left, accompanying the attention map by users quite well in its own right.

(Here instead was considered a version that was probably the correct one in the exam, given the “old-looking” layout structure you will usually find here)

The website presents with the logo of website on top, clearly displaying the “who”, also displaying the physical location of the website subject. It already displays a number of things a map aiding the user to get more context of what axis, also guided by a good number of links present in the left section of the main page, giving the user many choices and in a good position, considering it’s left and it’s the position the users favors the most. Overall, many choices are presented to the user, actually in a fairly simple way. Also, to appreciate the fact it offers multiple versions of the website, both in languages and presentation.

The navigation seems quite easy (actually, going into internal pages shows there is a breadcrumb, we can’t say that considering he gives us a static image) and also is guided towards more links, considering the grid layout of images present in the homepage, which seem all clickable and actually invite the user to other goals and means. There are actually multiple things to click to try to get new information and presenting content in grid might aid visual browsing and make the user get a better idea

The “how” axis seems quite well respected, at least on the offering of “where to go next” for the user; we can’t say the same for the search though, considering it seems quite limited on the number of characters employed, which is not good. To get the grasp of main content one has to go deeper in navigation, seeing videos, photos or reading; while the layout is colorful, it doesn’t have usability-first in its thinking, sadly.

The “when” axis of information is not considered, given there are not clear ways on how to get news from the site. Overall, the page seems pretty compact and seems to not require that many scrolls. The text itself is not really readable or coherent; different ways of formatting are present, but the font is quite small.

Written by Gabriel R.

Carnevale Olbiese

(Considering exam date, I took a 2016 version, available here – now in 2024 site is not even available anymore)

The site already tells us all we need to know about from the name: it represents the Carnival in Olbia (only downside, it's in Italian that this is understandable, but it was when the site was in Italian, so no problem anyway). Luckily on the menu there is a "Contatti/Contacts" to possibly get more info about this. From other perspectives, already the who axis is not clear: there is no clear logo on the top left and can get a grasp on who is behind the site reading the blurb present as subtitle on top of website, so an association of friends.

To accompany navigations, two menus are present, one in the top left giving some redirections to other content of the website (good position, it's where attention map talking, the user looks the most, but should be top-left to be even better), some content on center and a column top right about a gallery and news of websites. So here, the "when" axis is respected in full: press releases are available just entering the website. The "what" axis seems to be not so well treated: there are some links to give user more context, but the center of website is taken by lottery numbers, mixing both Sardinian dialect and three different colors with two different fonts: so here text should be the main focus, but it's done in a fairly distracting way, not clearing the context of what the site should be about.

It's a carnival so the main banner (not clickable, one can see that) reflects this (only in meaning, not in action), but one has to revolve on internal links on top-left menu to get other information, otherwise he's completely lost looking elsewhere. This has two consequences: the "how" axis revolves again on the same set of links, because also there is no search functionality, at least from what is visible here. Considering users have limited time, this forces them to spend more times to try to understand something, possibly wasting clicks (gambling clicks). Links seem fairly descriptive on what action they should do and the page by itself seems pretty "skinny" – few links/few scrolling.

This site seems simple to execute but fails on many aspects, hence the "why" axis, why to stay here. Information seems pretty disorganized and just there for the sake of being there. Redesign the website here should consider bringing some reasons to stay and visit the site, possibly encouraging scrolling and get a clear grasp of the main concept presented, for example redesigning the layout in a way content is present clearly without the need of scrolling and aesthetically pleasing.

Written by Gabriel R.

Musei Civici Bologna

(Website reference: http://www.museibologna.it/)

(The layout, 2024 here, seems pretty horrendous/obsolete, so gonna analyze that)

The website immediately communicates by its name what it's all about, which is quite good. Already, the layout is particularly disorganized: the site tells us two times who is behind the website and a huge banner is taken only to display the bigger logo of the organization behind these museums, while you can see a typical logo with a little blurb on it on top left. A good point is the presence of an "About Us" link, even separated graphically with a border effect on the menu on the left, which might give more info. It seems like they particularly care for us to know who they are.

The context seems pretty scarce: from an information point of view, there is no suggested or interesting content to see or consider here, given there are only images and no text accompanying the vision of the user, forcing the scrolling downwards to actually see something. We can immediately see there are some images, with the hopes they are clickable "to see something more interesting" (gambling clicks). There are news appearing, as it seems, on the website, going downwards, so the "when" axis is good. Like before, there's also a link separated with border effect specific about "News"; one could also get newsletter and press releases. There are also social links in case, which is a nice addition.

The menu is positioned, as happened quite often in old layouts, on the left side of the page, attracting focus because it also has a different formatting, making the user focus quite clearly on it (all grey with links). There are some links though which are "taken apart" from the other ones, which seem to attract user attention more; this can be good, but many don't offer the "normal" user no additional useful information (say, accessibility, transparence, trust for contemporary art). It would have been nicer to use a coherent formatting, possibly just blocks of layout separated with logical coherence (like: who are we and press releases/when then another block about the rest, all merged in color) and also a more readable font, possibly not uppercase. Consider also there are already 4 different fonts in a page, which is quite a lot.

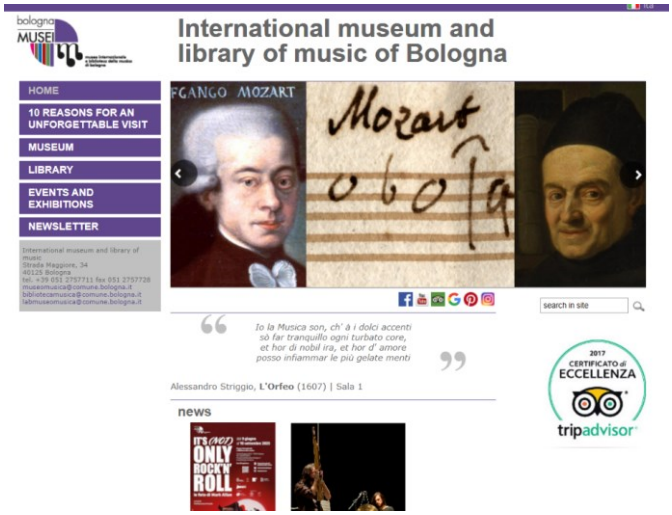
There happens to be a calendar, which might be to pay attention to particular museum events; in any case, its usage without any context from the start might puzzle users, which will never click on that considering the attention map is all taken by the links, which in this page are the only interesting part to look at and to understand "why" someone would have come here. On this aspect, consider the

Written by Gabriel R.

presence of a search bar, which respects the classical “Google-like” layout. It’s nice to have this option, but sadly it seems to have not even 20/30 characters, so users might not even see what they are querying or typing. One thing to note seems like the presence, on internal pages, of a breadcrumb to make users understand where they are; even if it was present, it definitely would be small because of all the space taken by the big logo, which is not good.

Overall, the site is pretty empty without clicking on additional links and this page does not give enough info to be useful.

Version 2



In this version of the same website, it seems like they did things better; the giant logo has been replaced by a carousel of images, with the hopes of those being clickable to help the user go next. The menu has been redesigned to be much more compact and even all separated voices; this is quite good, as it seems they are some “about us” kind-of links (museum, library), some “what” links (events/reasons for a visit) and some news even (newsletter); on this aspect, the news now are in the “safe zone” (immediately visible when getting into the page), quite better to get a grasp of the page content easily and have additional useful information.

About what the page offers, there’s a quote In the middle of the page which, even if poetic or just some fancy content put there for layout purposes, doesn’t give any useful/additional information to the user. Infact, if the user wants to know more about this site, he definitely has to click on the top-left menu (good positioning attention-wise); also, the who axis is perfectly respected this time around – logo top-left with two blurbs) then the title – actually giving some useful information, additional to the link of the site. Social-wise, there are many more links this time and also a TripAdvisor banner on the right as a “quality certificate”; again, there is already an icon for social links, so this banner takes useful space basically for no reason.

A good aspect is the presence of the contact information top-left; unusual choice, because this specific info usually lies inside the footer of websites and also it might be in a contact/About Us page. Again, to get more information on where to go next, consider the search bar, always small accompanying just a selected set of characters, not so good in the long run.

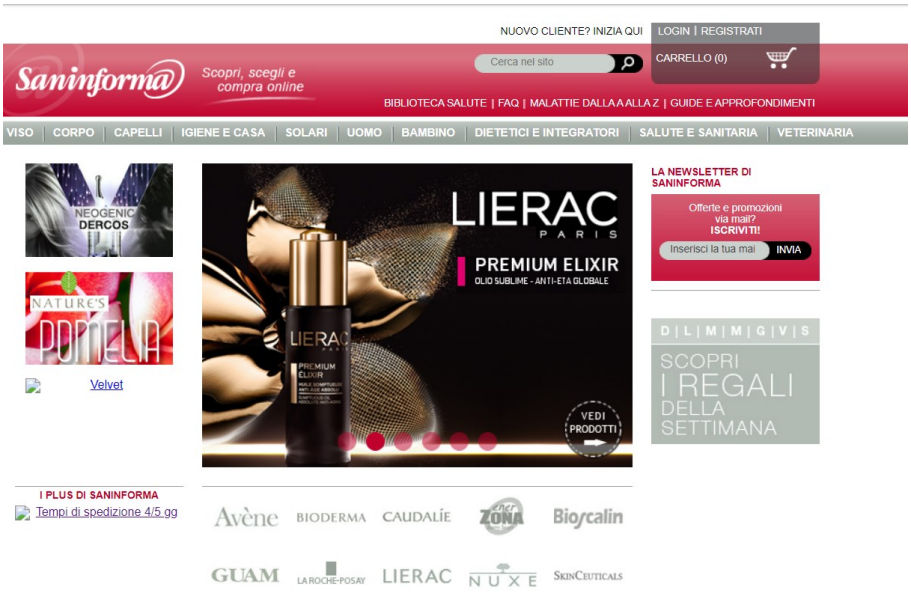
Overall, again, the site is better than the version before, but the “why” has to be satisfied only by clicking additional links of the menu; also this time on the portfolio, which has a purpose overall, interesting and useful. It’s better but it could be even better giving something useful for the user to stay inside of here.

Written by Gabriel R.

Saninforma

(Website reference: https://www.saninforma.it/)

(Also here, we analyze for the sake of presentation, an archived version of circa 2015, considering the layout the professor usually employs)



This website contains a word pun in its name in italian, so it plays around a slogan. Also, there is a real slogan right beside the logo: empty and adds nothing new to the table. The page presents many elements, and we will analyze such in detail.

Already, there is a newsletter asking the user about registration and personal data to send out a newsletter, which is not a good thing. The “what” axis depends on the navigation links and also the banners present in site; a carousel of images, which should be clickable redirecting to a product of the site and other banners. Basically, they all represent ads, specifically the left ones, which are positioned inside the most-focused zone visible by the user, a good thing. Also, they do not have any border to separate them from the layout, so it’s fairly good as an approach.

Still, they do not encompass text as the main content, because we basically only have a carousel of images and a set of sponsors, so ads should ideally accompany the main content of page, which should be text, but is not present. In the end, considering the nature of site, it’s not terrible as a choice, but a grid of products, possibly four by scroll with a small blurb describing them would be more enticing while not being too heavy for a browsing user. Already, there are at least 4/5 different fonts employed; even if similar graphically, it can become heavy. Possibly, a bigger font would better suit this kind of browsing (with some resizing options)

The links placement is horizontal and works well, but already, in the red overlaying bar, we see some fixed links always there to give FAQ and additional information; understandable choice, but this can overload the user with options. One could simply put a FAQ section appropriately designed inside the grey menu so to make layout cleaner and more minimal. The “how” axis revolves around the said links and the search option, which as many other sites, seems limited on characters.

Written by Gabriel R.

Here, we consider the presence of login/register/shopping cart, which is interesting and a bit non-standard. A good placement would be on top-right, which many sites do to accompany these specific actions. The optional “when” axis is not present but in the case of site of this kind is not particularly important; by looking at the site, we would probably get a glimpse of who is Saninforma just by going to the footer or something like that, no elements are present to give additional information.

The site overall is pretty static and it’s not terrible information-wise, it does its job fairly well, but could be better with the discussed adjustments.

Written by Gabriel R.

Dolomiti.com

(Website reference: https://web.archive.org/web/20190111112057/http://dolomiti.com/)

The website already tells us quite clearly the context and the domain name is pretty good, both in domain choice and name. The website gives an idea on who is behind the website, probably an organization curating the tourism in Dolomiti, but is not totally clear; anyway, the logo is present on top-left of site. Probably going to this page footer, like happens other times, may help to retrieve more.

The site offers some elements in grid in its center, with images which seem clickable, but I think they are not, so the user visual metaphor is probably betrayed; the links however seem descriptive. Also, we notice two sections of said website are like a “coming soon”, so they do not provide any informative value to the user. On the what and why axis, the site seems quite equipped, given it offers a series of links (mainly on top-left, which informatively and attention-wise is not bad) and also a list of other websites on bottom-left, which also give traffic to other sites.

This seems like the Jakob law principle done wrong: here we are inviting the user to directly go away from the site, even in multiple occasions. In this case, definitely we are talking about partner sites, but it’s a factor to consider. As happens on many sites, the “when” axis is not considered at all, there are no news to reach or read. The page seems pretty compact, so no useless scrolling is required. There are blurb over grid content, to actually inform the user or where to go and understand things about the page. This is somewhat fat, considering the number of accessory elements employed just to display a few links. The menus are linear and pretty usable in their own right.

Written by Gabriel R.

Carabinieri

(Website reference: https://web.archive.org/web/20190205145407/http://www.carabinieri.it/)

The name itself it's quite self-explanatory: a police force letting us now it can help us. In this case, the slogan might even welcome the user giving a sense of safety. In the nature of said website, it's simple, short, effective in short memory, so not bad.

There are some comments to look upon: the design by itself is pretty minimal, so the way to reach information would be only by going inside the tabs present. By themselves, they are not bad: they have a coherent layout, a small icon related to the context and even we can see indicated in red the tab of website section we are supposed to be in. This is like a "visual breadcrumb", so a metaphor helping the user knowing where it is. Another welcome choice.

We know what this site represents thanks to tabs like "Arma" and "Contatti" which can make us know more (also to get help fairly quickly given the context); not bad again. There is also a "social" aspect with the "Community" tab; this can encourage user to login and get more information. On this, the login aspect is quite old and quite shabby; the links hardly visible higher and lower than the search bar, quite small and with search icon fairly hidden in website. While the search bar, along with links, accompanies the "how" axis, the login links have no purpose in being there like this.

Possibly the content here is already images; the "where" axis is taking up (or better, should be, given the actual version of this site when there is the white space accompanies news) the whole section given press releases, environment initiatives and events. So the axis is quite rich in this case; the site offers it all as the main content but maybe it's something meant more to internal people and shouldn't be forced all page inside user throats like this: so I think it's a debatable choice. There is also a banner, telling us that this site won an award; interesting for something like a second, again, completely useless information-wise and just empty as a slogan.

Font-wise, the site is not terrible: it uses three fonts fairly similar to each other but should be as big as the one present in tabs, which is quite good in my opinion. A good and simple choice would be to adapt that font, at least as characters if not size to influence vision on what to click first, everywhere just to keep visual coherence.

Written by Gabriel R.

The user should stay inside site to get more information about contests and to get help if needed via ease, but apart from that, there are no particular elements which trick user staying inside website because of reasons. Informatively, it's clear one would get here by deep linking, so via search engines, because even if clearer than many websites, it's pretty empty and very much devoted completely on itself, so the whole Arma, but not other things.

Written by Gabriel R.

La Versilia

(Website reference: https://web.archive.org/web/20150205232528/http://laversilia.it/)

To quote what James Rolfe would say: What were they thinking? The site is a mishmash of so many chaotic things it drives away people attention within milliseconds.

Let's start by the safe zone, which would be nice if only it was empty from all of those links; a grid layout composed of four different tabs all completely different from each other, separated by other sponsor hotels tabs not even coherent in size. The site clearly thought spamming the Versilia term would be good for SEO purposes, but drives people away, given the menu is a tab inside a tab offering a lot of empty links. It would be much better to have at least 10 less links and without the word to give time to breathe for the user.

Let's not talk about the "wonderful" WordArt banner, which defined as atrocious would definitely be the best compliment; completely fancy, unnecessary and just void anyway from any possible point of view. The only part useful if implemented well would be the second banner on its right and the one left on second row; they talk about offers and restaurants. A good way would be replacing this with some search bar and filters to allow a unique way to look for both of them (spoiler: this is what the actual version of the site did, luckily so). As a positive, the menu bar on top is pretty informative, but if one user wants to get information, he has to do gymkhanas all the way.

No search bar is a terrible choice, but coherent to a terrible website of this kind. Don't worry, they got us covered letting us know who they are; there is a banner dedicated to that on the right in second row or also the very old log on top with a completely unnecessary and annoying banner of top. From the look of page, it requires even scrolling with this much of information they throw at us with basically no context apart from annoying us. Contrast wise, different fonts with different colors and pretty small without resizing and I would also bet this site would force some kind of horizontal scrolling given it's brick-layout.

In conclusion, the site is terrible and the best way to use it is to go into another website different from this one, probably. A redesign of it should consider all the points previously discussed.

Written by Gabriel R.

NOI Padova

(Website reference: https://noipadova.it/)

(I consider here a 2018 Web Archived version; I won't put the link because it was painfully slow to load, and to have it complete I had to wait for a non-broken version for a while)

The site presented is interesting, both in color layout and navigation. The "who" axis is clearly displayed: site name and logo with little blurb, actually informative, top-left and also an "About Us" link in the menu left. Also consider the social links on the top-right. Good job. Basically, given it's an association, the whole site revolves around the "when" axis to be the main content, so to be the "what" axis. This time around, the "when" is quite riche: main content and two dedicated to menu sections.

The news are static, so each one tries to give some blurb, a category a title, but one has to click the link (there is no "continue" after suspension marks) in order to possibly go ahead and read more. Given the content mix, one would appreciate more how to get some benefits immediately, but here giving the news upfront seems like the site is meant for an internal audience, who already know the content of the whole association. This definitely is the case here; we have a "Link Amici" – "Friends link" on top right, before login option and "VP users", which "should" represent some higher-level user. Again, this is understandable, but these options should have two sections when designing login, putting in the usual top-right registering and login for all users (in the current website version, this happens).

The color choice for the main browsing menu, the top one after the Padua image banner, isn't very visible, given it's orange on orange. The main content, at least, with light-blue and orange itself both uppercase and lowercase, it's pretty coherent. Definitely one would avoid using the top menu when it has the "rainbow" one with all palettes of colors top-left, even with a mascotte of sorts. This in itself is pretty fancy looking and distracting from the other site content; even all colorful pins inside content; nice looking, help differentiate the news, but are there just for mixing things up. One should prefer some content there, some more minimal approach and just using an orange-blue palette also for the top-left menu to achieve coherence.

Written by Gabriel R.

The ”why” here seems to consider, as said, only people who already know the association; one should focus more on putting projects and things here on the page to possibly give interest to the normal user, considering he probably has no clue what this is like even “VP users”/”Friends”). The search functionality is present (a static/internal one), apart from the main links, to get something from the “how” perspective; quite small, no search button, little characters employed. Also, the banner “NOI Hub” should be probably an external site, but it’s there with no context in its presence.

Attention wise, it’s good structure (or at least the base is here), but content misses the mark, mainly, to properly redirect user navigation towards something useful. One has to go deeper in navigation in hopes to know more about “what” and “why” to stay here.

Written by Gabriel R.

Magnetic Island Website

(An old version of the website https://www.thisismagneticisland.com.au/; the actual version uses a hand-written version to symbolize "tradition" if you click here – amazingly bad as a choice I'd say font-wise. Content-wise, it's not pretty bad)

Getting into the site, we immediately notice some not-standard things; the site is meant to represent the Magnetic Island, but the first thing to see apart from menu is an ad to fly to said destination – amazing stuff. On top-left we have a static image of sorts, and a menu displayed sequentially with elements one after the other, even with empty voices and with the home page link called – for the sake of being alternative – "front page". Again, amazing stuff – non-standard and simply puzzling.

The content of the page, after the title, follows a "slogan" of sorts/banner useful for nothing, which should entice the user to get some grasp about the "why" and the "what". The browsing is naturally guided towards the clicking of grid elements which follow a Q&A/W axes structure, meant to give something for the user to see. If one gets here, has to make effort to see more information because the homepage already gives some empty content and wants us to know more and see other things.

To get to other things, one has to revolve around these ones; no search functionality is visible, one could click on the main elements or the links on the side, with a layout not too visible and hardly interesting in color choice. We have at least, considering the ones in images, 6/7 different fonts employed, symbolizing overall "fanciness" compared to usefulness in graphical design. If one gets here, to come back home, has to click the "front-page" button, or just uses the back button multiple times and goes to other things.

Talking about the ad, it tries to integrate with content in the second-best position of attention for an ad, but uses too many fonts and colors, possibly distracting the user from the main content. In this case, the ad is content, because the flight is required to get to the island and also tells us some little slogan to give us information about this one. Still, probably, make it less of an ad and more text-based, possibly less fancy-looking and place it in position of the empty slogan inside of page we saw before.

Commenting more on the menu structure, to say it's interestingly bad it's a compliment; as said before, empty menu voices, but the button "main menu" has no purpose there. It's like the "click here"; no additional value is provided on an information level and as many other things it's just there for the sake of being there.

Written by Gabriel R.

While the site tries to be interesting on paper, it fails on little things which slowly become big when gathering everything: coherence, menu presentation, content presentation and just overall color/content choice. Given it's supposed to be a visit card of the island, it completely misses the point of attracting interest, given browsing should be natural and not a burden because of surmounting problems.

Written by Gabriel R.

Arrestling

(From what I got looking for FinalOps on Google, I found the site of reference for this one was: https://arrestling.com/)

The site presented puts the name of site inside the main banner, not clearly making us understand who is behind the website, given its non-standard structure both of site itself and its menu. Overall, the structure is pretty compact, but it's clearly disorganized.

The main menu usually accompanies on the left side of the title, when positioned top-left with a logo with a blurb or the title of website alone, but here is presented with four main sections, where each one is then "exploded" into multiple sublevels, as you can see above. Usually, that section is dedicated to dropdown when hovering on a voice, but given we miss the style completely into this page, this is the main reason.

The whole site is in uppercase; the font is at least consistent, apart from the ad, which is positioned inside the content (not separated from it at least), but in a bad position for the user attention and clearly activating the natural zapping effect of the normal user there. Also, it's completely colorful and overall distracts too much from the rest of the content. The links make us understand where to go to and the "what" axis is at least covered with content, directly offered also by the main 4 banners after the menu. Their structure is pretty harsh and disorganized, given their lack of style, coherence with website and just overall pointless nature of the images, put there just for graphical reasons and not as context.

On the "where" axis, it seems like the only way to come back home is just clicking the link you see in the menu, which is different from clicking the site name. The user will definitely prefer clicking the back button to move such. As of now, we have no search option employed; to move around, only clicking links is needed. Also, no news are seen as of here, but it's not mandatory given the quality of the information we see here. Still, the site updating here is definitely not the priority.

The "why" axis seems pretty condensed; a user would stay here to get information about training groups, corporate and law enforcement, so one has to go deeper in navigation to try to get some more content out of this, given one would arrive here via deep linking only if looking for the terms specified, but has to get more time trying to navigate a not-so-ordinary and old layout to possibly find something.

Written by Gabriel R.

While the site is simple, it could be better by simply reorganizing its information according to a normal “logo-menu-content” structure, while also putting a list of condensed links on top or even left. Inside the page, put some simple content, possibly some images with blurbs or some text required to get more information about upcoming links. Put some more style and some coherence overall and it can become functional.

(In current version of the site, it’s better in menus, but we have a static carousel of images and four videos while scrolling the homepage; not very useful at all I would say).

Written by Gabriel R.