T.M. Basyuk, A.S.Vasulyk
Lviv Polytechnic National University, ISM department

RANKING WEB SITES ON THE INTERNET

In the article are described designing features of ranking websites on the Internet, are held the analysis of known systems of resources promotion and are shown internal and external factors of ranking. Possible schemes of organization of the Internet resources with link ranking use are analysed and suggested. The article describes the main problems that appears during optimization of pages for the set of requests and gives the scheme of distribution of keywords on resource pages.
Keywords - rankings, internet resource, relevance, optimization.

Introduction. General formulation of the problem

Every day the problem of analysing and ranking internet resources draws attention of more and more researchers and experts in the field of The Internet Technologies. Since there appears a considerable number of new projects which are interesting and perspective, but remain unnoticed among a wide variety of already existing web sites. Because of that, the important characteristic of generated resources, at first, should be determination of target audience and regional features, as long as creatinating websites only for people can make their indexation and interpretation by searching works, so it can take a lot of time for website to develop and to reach its purpouse.

Nowadays the development of internet resources should be done in two ways: on the target user and on the searching robots. Mentioned feature promoted appearance of the concept of search engine optimization. Under search engine optimization (SEO) we understand complex of measures to raise the site positions in the results that search engines give for users certain queries. Namely, the higher is the position of site in search results, the more interested users move up to it from search engines. Analyzing the efficiency of search optimization the cost of a target user is estimated considering the time we need to take site to the indicated positions and conversions of the site which they enter [1,5].

Conection of the depicted problem with the important scientific and practical tasks

There is a chain of companies (Webeffector, Megaindex, Seozavr, Trustlink, Upominator) which are responsible for promoting internet resourses. The result of their work is web-site ranging by increasing its rating in popular search engins: Yandex, Google, Mail.ru, Tut.by [3].

It is complicated continuous process, during which methods and algorithms of optimization and search are widely used. Implementation of this task will make it possible to formulate special features of designing system of resource ranging and will give means of adaptation of famous methods and algoritms to the problems of decision support in particular field.

Analysis of recent research and publications

The problem of ranking websites appeared recently [9,10], therefore its solution is still scantily explored in general scientific stream [3,6]. Considering that, analysis of latest researches was conducted on the basis of known resources which provide means of analyzing and determination of the certain websites popularity. According to the Computer World Ukraine magazine rating, among the common systems that functionate on the software market of this class we can distinguish: site analysis system (cy-pr.com), seo-tools (seo-tools.com.ua), tools for web Wizard (pr-cy.ru), setlinks (http://www.setlinks.ru/ stat/) and site-auditor (http://www.site-auditor.ru/) [12].

Site analysis system (cy-pr.com) - is considered to be the leader in analysis of internet resource options on the software market. It displays information about the resource in two blocks: the first contains basic information about the site name, keywords, IP address and server location. The second

block shows information about a web site which is contained in search engines databases: Yandex indicators TCI, Google PR, indexed pages, availability in directories and possible profit from links selling. The disadvantages of this system are commercial use and the lack of any information about possible ways (recommendations) of ranking internet resources after the analysis process.

SEO-TOOLS (seo-tools.com.ua) - this system is very similar to the previous one for its functionality. Output of results is realized in two blocks, the first of which also contains basic information about the website: its name, keywords, description and IP address. The second block of results contains: key indicators of website (TIC Yandex, Google PR, Alexa Traffic Rank); external links, information about presence of the site in directories, indexed pages, links on the other resources. Like the previous system, SEO-TOOLS makes analysis of site parameters and does not provide recommendations for their modification to increase the rating of internet resource.

Tools for web Wizard (pr-cy.ru) - analyzes the content of corresponding online resource. Unlike the previous systems, its functionality is considerably lower. The peculiarity of the system is an output of information about TIC Yandex, Yandex Rank, Google PR and Alex Rank. Information from the last resource is displayed as a graph that shows the most popular keywords in different search engines with positions in them and direct links. The disadvantage of this system, like in the previous ones, is lack of recommendations on how to promote internet resource, and the specialty is an availability of means of keywords analisys and their position in search engines.

Setlinks (http://www.setlinks.ru/stat/) - this tool provides means of express analysis of internet resources that affect their position in rating in search engines. It is possible to make analysis by keywords, internal and external links, and to identify the closest competitors. Service provides ways of finding: site authority indicators in accordance with their raintining in Yandex TIC, PR, Alexa Rank search engines; the keyword density on the page; the list of words by which the site is in the top ten given by Yandex, Google, Rambler; the list of the closest competitors; the quick access to links on pages of main search engines (Yandex, Google, Rambler, Bing, Yahoo, Mail).

Site-Auditor (http://www.site-auditor.ru/) - the utility which accumulates necessary data for ranking site in search engines. Almost all necessary data from the major search services that are used in Russian and Ukrainian parts of the Internet, namely, Yandex, Rambler, Aport, as well as the most popular international service, such as Google and Yahoo. The specialty of the system is that in case of counters presence there is conducted gathering of their indications for the last week. At the moment, the system detects the availability of the folowing, the most popular statistics counters: Rambler Top100; Mail.ru; LiveInternet (RAX); SpyLog; HotLog. The utility provides means of determination of site position in search results in systems Yandex, Rambler and Google by selected request list. All collected data is stored on a user's computer where the ongoing changes are dynamically monitored. There is information line at the bottom of the utility window that displays RSS news line.

So, in a result of the analysis it was found that famous software does not meet the realisation of resource rankings requirements: they have lack of recommendations on how to promote internet resource and commercial use (site analysis system (cy-pr.com), tools for web Wizard (pr-cy.ru), site-auditor (http://www.site-auditor.ru/)) makes the task of developing the web sites promoting system topical that will ensure operative analysis of resources and will give recommendations to improve their rating.

The main issues of research and their significance

Purpose of research is to find specialities of projection of ranking websites system, which will provide means for conducting complex resource analysis and, depending on its results, will give recommendations on its improvement in order to promote it in search engine raiting. To achieve the goal it is necessary to solve the following main tasks: to make analysis of known ranking technologies; to adapt the chosen approach to the settled task of building the analysis system and promotion of web sites; to give recommendations about creating of internet resource depending on keywords. The projected system should solve the following problems:

• collection of data about internet resources and their systematization;

• output of extra knowledge from already existent in the database;

• searching of additional information about the parameters we analyze;

• output of recommendations for resource promotion.

After website optimization, its owner can expect improvement of its positions in search engines, which will increase the number of visitors, what is crucial on the Internet. All results of the fulfilled work solve the actual issue of creation ranking websites system.

The main research results

Location of Internet resources in the top lines of search engines can be achieved in different ways: both legal and illegal. Considering that there are different types of websites optimization: white, gray and black. The white optimization is a work with resource without use of officially banned methods of resourse promoting - without affecting search algorithms. It includes work with internet resource, namely, with the internal navigation and content; with the external site environment, that is, promoting by making reviews, press releases and signing up for partnership programs, etc. [2.11].

To gray search optimization we include means of adding a large number of keywords into text on a page. Mantioned manipulation often affects the resource readability, for example: "Salt is salty". Herewith, firstly, optimization lies in the selection of key requests for particular web page, in finding of needed frequency of keywords in it, and then in formulation of sentences and phrases that contain a certain number of key requests in different cases with different forms of verbs. In this case, the task of person who makes resource promotion is to write original texts in the way that such optimization was less noticeable for a "live" reader and moderator of search engine. Another example of gray optimization is a doorway without redirect, which means that in case of a contact with a doorway there is no automatic redirection to the resource that is promoted [4].

To black optimization we include all methods which contradict the rules of search engines and, as a result, entail blockage of the target project. Among them we can highlight the following: use of doorway (pages and resources created specially for search engine robots, often with a large number of keywords on a page), cloaking (for user is shown one page that is easy to read, and for the search robot – another, optimized for any requests), the use of hidden text on the site pages, the use of "single-pixel links."

Another thing which affects the position of Internet resource in the global network is ranking: internal and external (Pic. 1).
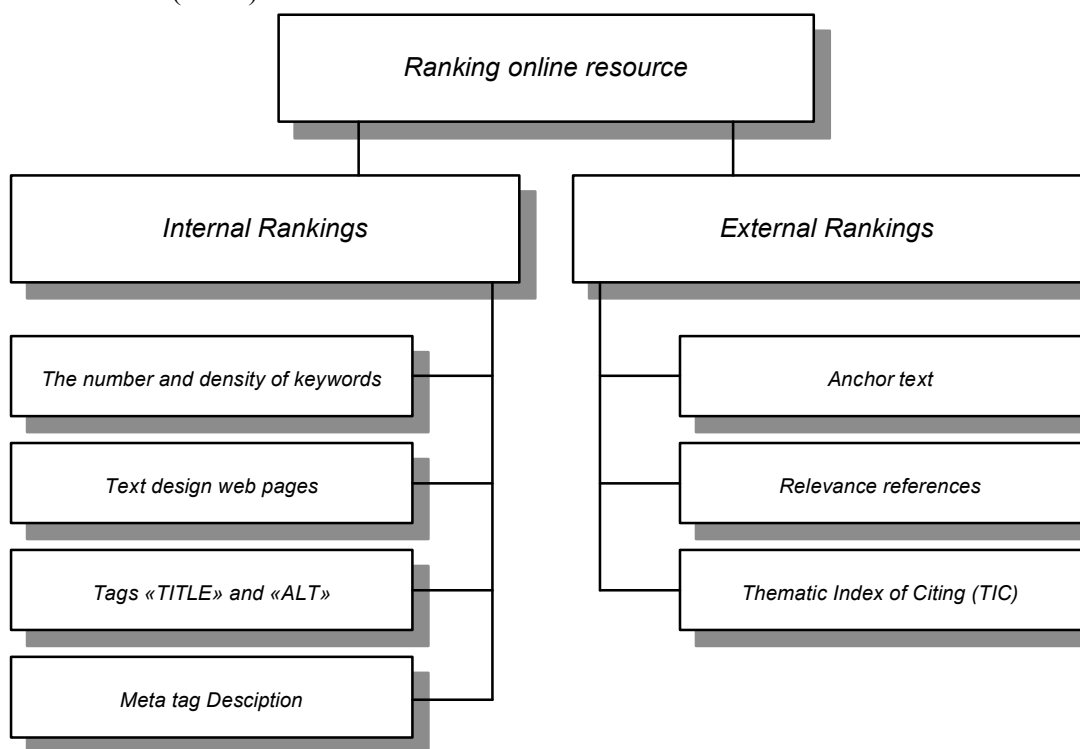
Pic.1. Internal and external factors of online resource ranking

Internal ranking factors – those ones which are controlled by the website administrator (text, design, etc). The main internal factors that influence resource position are: the number of keywords and their density, text design of the page and meta tags (Pic. 1). The main elements which define the internal ranking are [8]:

– *the number and density of keywords*. In accordance with this parameter keywords (phrases) should occur in the text, at least 3-4 times. The upper limit depends on the total amount of pages - the larger the total amount is, the more repeats you can do. The density of the keyword on the page shows the relative frequency of word repetition in the text and is measured in percentages. For example, if a given word appears five times on page with one hundred words, then its density is equal to 5%. Low density (2-3%) leads to a leveling of a keyword by search engine, too high density (10-15%) can activate the spam filter of system engine. The optimal keyword density is 5-7%;

–*text design of web pages*. Search engines give higher positions to the sites with united information content. Optimal are pages which contain approximately 500-3000 of words. Larger amount of the text on a page increases the "visibility of the page" in the search engines due to rare or random search phrases;

– *tags «TITLE» and «ALT»* - the most important tags that are actually a visiting card of a page. Tag «TITLE» should be informative and appealing, and its size usually is limited to 50-80 symbols. Any image on the page has a special attribute "alt text", which is set in the tag «ALT». This text is displayed on the screen in case it is imposible to show a picture. A speciality of this tag is that existing text is to be ranking in case of relevant links, which is an additional tool in resource promotion;

– *meta tag Desciption* – is for giving a page description. It does not affect ranking, but is very important because search engines display information from this tag in search results. Therefore competent compilation meta tag Description is important in terms of attracting visitors to a particular web resource. The description should be brief but informative and attractive, contain keywords that are specific for this resource.

Along with internal factors significant role in internet resource promotion play external factors which include:

– *anchor text (link text)* - the text that is positioned between the tags «A» and «/ A». If the link text contains keywords, than the search engine sees it as an additional and very important recommendation, evidence that the site really contains valuable information relevant to the topic of the search query;

– *Relevancy of references* – affects the overall information content of the page. For example, in case of promotion of the computer store resource, links to the site of processors production will be more appropriate than a similar link to the site of medicine or herbs. An important feature of this is the accounting of the external links to the site - the more links are, the higher interest of visitors is. It is considered thet if other online resources have links to the estimated resource - it is quite good, and because of this the search engine gives it more importance;

– *thematic index of citing (TIC)* – is calculated for the whole site and shows the authority of the resource regarding other thematically similar sources. TIC is used for ranking sites in the major directories. Besides the usual citation index, which is an absolute indicator (is some specific number), there is a term weighted citation index, that is a relative value, and it indicates popularity of the page towards popularity of other pages on the Internet.

In addition to these factors, the ranking of resources in a global network is realised by another set of factors that affects the resource structure. It contains elements [7]:

– *main page of te site* - optimizing homepage (domain name, index.html) to the most important phrase, we receive increasing of Internet resource rolevance;

– *number of pages*. General rule: the more - the better. Increase in the number of pages improves site "visibility" in the search engines. Besides, gradual addition of new information materials is perceived by search engines as a site development, that provides additional advantages in ranking;

– *navigation menu* – supports a transition between resource pages. An important condition of promoting is the use of keywords in the menus, what gives extra importance for pages to which the reference is. Using keywords in the HTML-File Page title positively affects site position in search results.

All things considered, we can say that for increasing the page rank we need to work to make bigger amount of network documents be linked to it. This can be done in various ways - by exchanging links with other sites, registration in catalogs and various thematic resources etc., the perfect way is to make the site so unique and interesting that owners of other resources wanted to put a link to it.

Link Ranking

One of the ways of ranking is search through link vocabulary, so-called link Ranking - the impact of links text on the page and its relevance to search request. In other words, if the words from the query meets in the text of link on the page, it increases its rolevance for this query. In entering this ranking parameter in search engine algorithm lies a general statement: if someone refers to a page with some text, then with high probability we can be confident that the information contained on the page matches the query, and the more such links are, the higher this probability is . And if the site that is referenced is popular with many users (is authoritative), then matching page content to link text is more likely - authoritative site does not contain low-quality content. Dependence authority indicators of the page is determined according to the formula [9].

$$L_a(x) = \sum_{i=1}^{n} La_i(x) * f(PR_i) \qquad (1)$$

where, PR_i - the authoroty index of the page i, is constant towards to the query x;

$f(PR_i)$ - function of authority indicator has a linear form, i.e. f (PRi) = k * PRi, where k - coefficient which takes into account the number of keywords and phrases;

$La_i(x)$ – relevance of the request x to the link from the page i. If there is no word from the query in link text, then $La_i(x) = 0$. The function $La_i(x)$ takes the maximum value if there is the exact occurrence of the searched phrase in the link text.

Thus, the magnitude of the effect from link ranking depends on three parameters: rolevance of link text, authority of the page and the source which is linked to. Often, during issuance in analyzing the site page it is difficult to isolate the impact of exact link ranking. However, this option is extremely important in case of the minimal influence of other factors. In Ukrainian Internet space the most widely used Internet resource is Google (about 82%), so the promotion of websites and their optimization is mainly based on Google Page Rank parameter [5].

Google Page Rank is calculated for each web page separately. The main task is to find the criterion that determines the importance of the page. In case of Page Rank as a criterion is used a theoretical attendance of the page, which is calculated as the probability of user presence on this website, while the sum of probabilities of all network websites is equal to 1, because the user is on some page anyway. For ranking Google search engine uses a so-called thematic Page Rank, ie one that only considers links from thematically related pages. Page Rank value is determined with using an add Google ToolBar, which reflects the value of Page Rank in the range from 0 to 10. In practice, Page Rank is used mainly for two purposes: for quick assessment of the site rating in search engines and assessment of competition in search query. For websites is adhered the following relationship: PR 4-5 - the most typical PR for most sites of the average development; PR 6 - developed site; PR 8, 9, 10 - found only on sites of large companies (Microsoft, Google, etc..). Values Page Rank is also used in links exchange to assess the quality of the proposed sharing page.

Another way to increase the resource rating is usage of internal links [8]. In general, internal links provide a means of redistribution of static weight (pagerank) between the website pages, thereby increasing their rank, what helps to increase relevance during search. The principle of static weight redistribution of the page is widely used during both the resource creating and its modification to improve the ranking level. Namely, during the "internal redirecting" from one page to another, static weight is transferred, yet not

complete but at a rate of 0.85 (attenuation coefficient is considered). All calculations in different circuits of internal redirecting are performed for several iterations, until the difference between adjacent calculations will not be small enough. Then it is considered that the static weight, which is transferred by reference in this scheme of redirecting, was calculated. In case of resource promotion a necessary condition is to increase the static weight of resource pages (articles). In general, external references in most projects are tabulated at homepage. Because of this, to increase the resource rating it is necessary to redistribute the static weight from the main page to internal ones, which contain articles. This scheme will generally take the form presented on Picture 2, where the static weight of resource pages is determined by Page Weight Online (http://www.page-weight.ru).
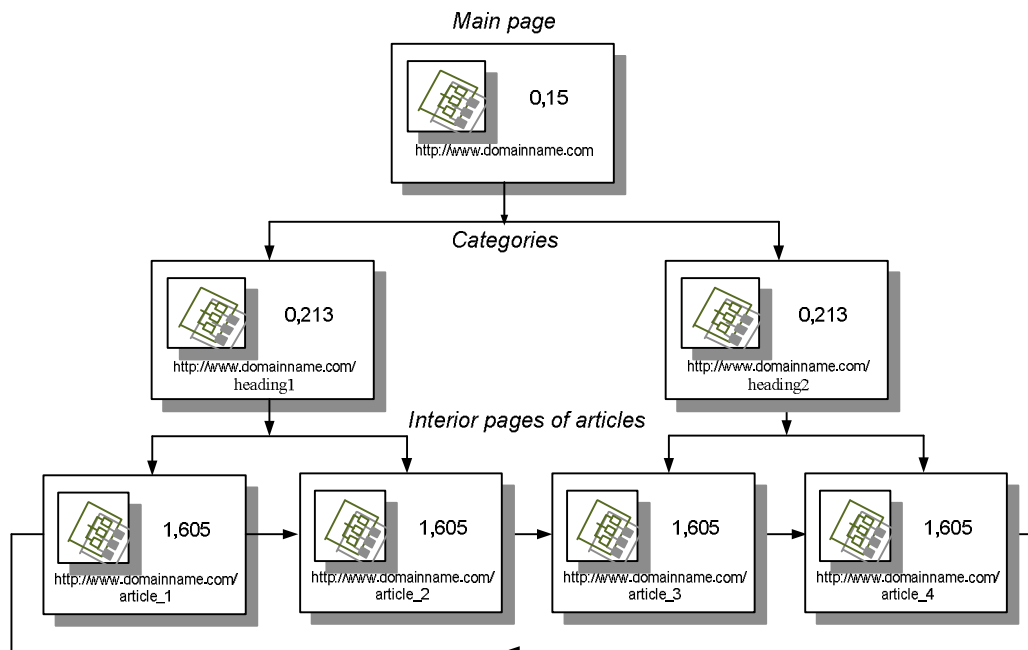

Pic.2. Scheme of redirecting of resource pages

As we can see from the picture, the arrow shows the direction of internal links, and static weight of the resource redistributed on pages with articles, what reduced the static weight of the homepage and its sections. Further increase in resource weight can be achieved by setting on the homepage of several external backlinks (Pic. 3).
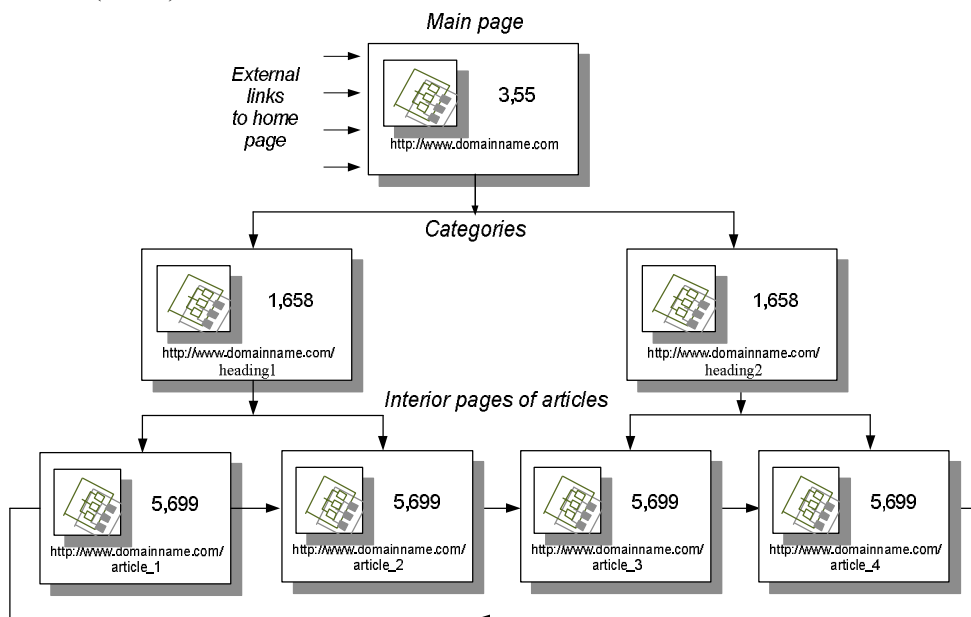
Pic.3. Installing of external links on the homepage of the resource

Looking over the above information, the best way to promote website with emphasis on the articles is the location of external backlinks not to homepage, but to the pages with relevant resources. In this case, the scheme takes the form (Pic. 4).
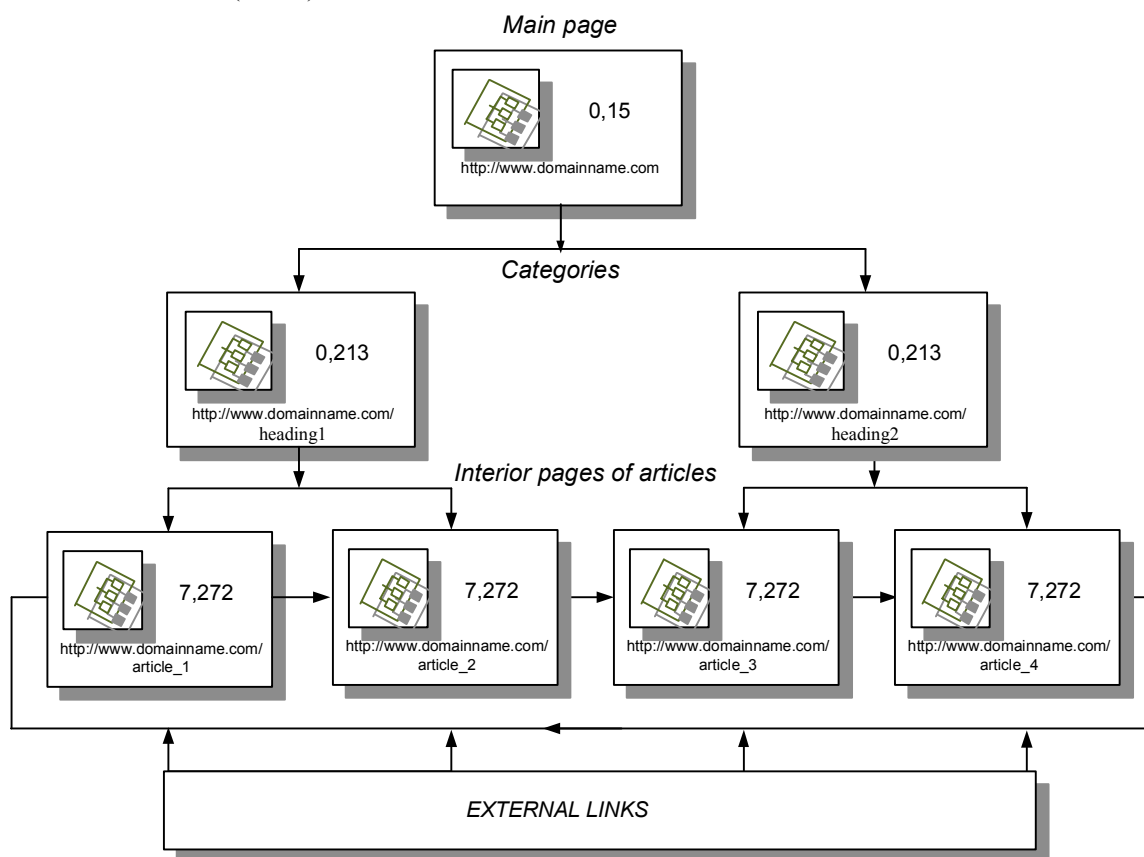
Pic.4. Installation of external links to pages with articles

Realisation of this scheme within a real project is done by placing a small block of links at the bottom of the page (prompts the user to material with similar subjects) and placing links in the text of articles, what improves transition and increases the number of pages viewed by one user. This figure of the review depth is taken into account by search engine while making a rating of exact resources.

All things considered, we can conclude that for improving resource ranking it is more effective to use organization of new links to pages with text than to search for the perfect set of values for the internal factors that is challenging task of multiobjective optimization [4].

Optimizing pages for multiple requests

Despite the relative simplicity, nowadays there is no single approach to determine methods for splitting keywords list into multiple separate pages or their grouping with further use. While grouping, creation of fewer number of pages, and, therefore, internal number of links, are made, but the disatvantage is reducing of links depth, and in reverse. Considering these features, there are two important and interrelated problems: the number of keywords on a page that have an impact on the visitor, and competition between the entered keywords [1,2].

Regarding the first problem, accoading to the statistics, while searching the information in Global network, users practically always are interested in an answer to a particular question. It is optimal if the answer is found on a separate page. The more diverse information is available on one page, the lower is visitor's conversion, and, as a consequence, the possibility of keeping potential buyer on the site. According to a media company AOL, conversion on a rate of 10% for the 10th search result is a better number than the conversion of 0.5% for the first position.

On the other hand, while splitting the list of keywords to separate pages, there comes a situation when in such words (phrases) basis remains unchanged. Picture 5 presents similar specific phrases that do not require clarification.
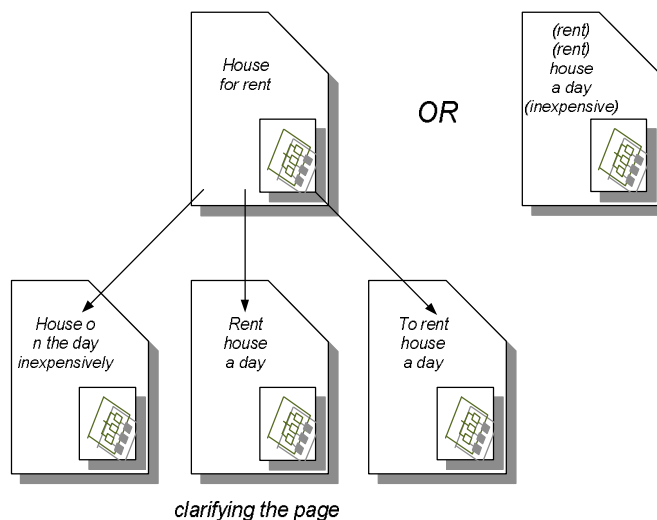
Pic.5. Typical circuit of spliting keywords to the resource pages

Using statistical services (e.g. wordstat.yandex.ru) it can be noted that for each of clarifying pages users make in avarage 100-300 requests per week. Given that, many web developers create for each of them their own page, which significantly reduces the quality of the resource and its content. Moreover, judging from the experience of Randfish (SEOmoz.org), these manipulations are negatively perceived by search engines and are rated as banned optimization with subsequent removal from the list of resources.

For the second problem, its feature is determination of the competition level in the theme of keywords. If competition is high, it is essential to create more pages, which are optimized for highly specialized queries with its simultaneous redirecting. And consequently, their static weight will increase due to other resource pages. Considering the described mechanisms, we managed to build a circuit of action for creating a resource with various keywords.
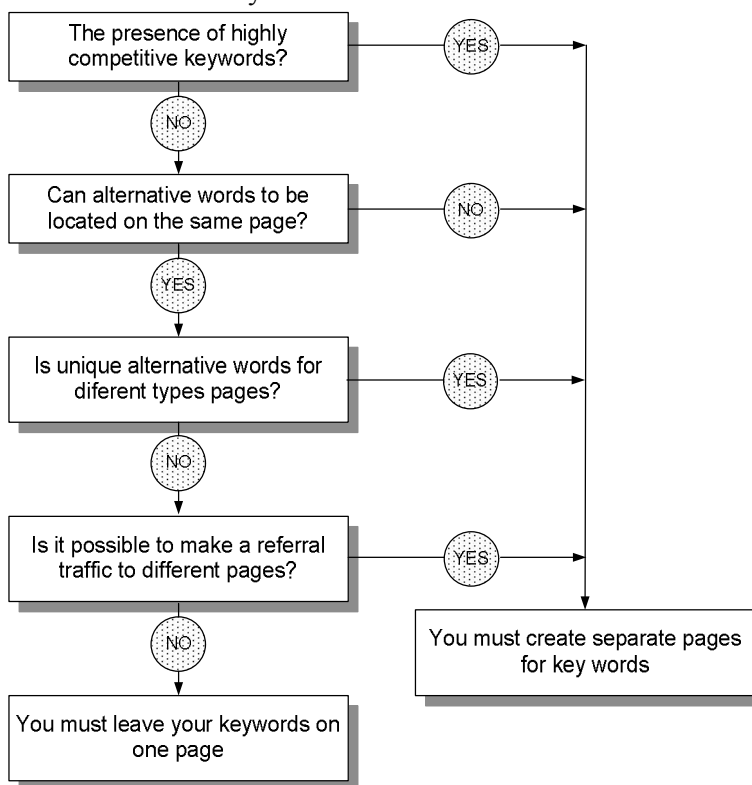
Pic. 6. Scheme of splitting process of keywords to the resource pages

Taking into consideration the above scheme (Pic. 6), in case of using highly competitive keywords it is appropriate to design separate pages with their further optimization, together with the use of internal and external links to certain keywords and phrases. In other cases, you must keep these words on one page of the resource.

Conclusion

As a result of the study it was found that for a successful process of the resource ranking necessary factors are: reliable informative of a website according to a theme, easy navigation, competitor analysis and optimization of website code. Moreover, a necessary condition is the location of backlinks on internal pages of a resource, what will increase their specific weight, and, therefore – their rating. Using of the described measures increases the efficiency of the resource promotion with its possible entry in search engines top list. Further research will be focused on designing of separate ranking system modules and verification of their work.

1. Ashmanov I.S. Advancing site in search engines / I.S. Ashmanov, A.A. Ivanov. – M.: Williams, 2007. – 304 p. 2.Bajkov V.D. Internet: information search and SEO / V.D. Bajkov. – SPb.:BHV, 2000. – 288 p. 3.Basyuk T.M. Principles of systems analysis and promotion of internet resources /T.M.Basyuk// Computer Science and Information Technology. – Lviv: National University "Lviv Polytechnic". – 2012.– № 784. – P.43-48. 4. Evdokimov N.V. Promotion website: A Practical Guide./ N.V.Evdokimov, I.V.Lebedinskiy — M.: Williams, 2011. — 288 p. 5. Zuev M. Website promotion in search engines./M. B. Zuev, P. A. Maurus – K.: Binom. Laboratory of knowledge, 2007. – 304p. 6. Christian D. Search Engine Optimization for ASP.NET / D. Christian – M.: Williams, 2008. – 400p. 7. Krokhina O.I first book SEO-copywriter. How to write a text for the search engines and users. / O. I. Krohina, M.N.Polosina - M.: "Infra-engineer", 2012. – 216 p. 8. Sevostianov I.O Search engine optimization. A practical guide to promote the site on the Internet. / I.O.Sevostyanov – SPb.: Peter, 2010. – 240 p. 9. Tero C. Visibility on the Internet. Search Engine Optimization. Edition / Sh.Terou - M.: Symbol, 2009. – 288p. 10. Enzh E. Art site promotion / E.Enzh - SPb.: BHV-SPB, 2011 – 592p. 11. Yakovlev A.A. SEO and website promotion: the basics, secrets, tricks. / AA. Yakovlev - St.: BHV-Petersburg, 2007. – 336 c. 12. Janowski M.V. System SEO-optimization / M.V.Yanovsky / / Computer World Ukraine. - By: Prepres. – 2011.– № 8(2011). – P.72-75.