

# Motorul de cautare Zamolxe

## 1. Continutul memoriului

Prezentul document este destinat analizei posibilitatii de implicare intr-un proiect de dezvoltare a unui motor de cautare romanesc si a unui brand de cautare si analiza a informatiilor web, pe baza unui prototip existent.

## 2. Privire de ansamblu

Intr-o lume in care volumul informatiei (in special cea electronica) creste exponential, iar nevoia de a filtra aceasta cantitate imensa de date este tot mai acuta, apare necesitatea unor motoare de cautare profesioniste, focalizate pe seturi specifice de documente, care sa furnizeze rezultate exclusiv pe criteriile de relevanta.

*Zamolxe* este un motor de cautare inteligent, bazat pe analiza semantica a documentelor in limba romana. Aflat momentan in faza de prototip, datorita rezultatelor mai mult decat promitatoare in domeniul relevantei, Zamolxe poate constitui o alternativa pe piata romaneasca, avand potentialul de a accesa intr-un numar teoretic nelimitat de domenii neoptimizate din punct de vedere al regasirii datelor.

## 3. Motoare de cautare

Un *motor de cautare* este un software dedicat regasirii rapide a informatiei, pornindu-se de la o serie de cuvinte cheie introduse (query/interogare). Pe baza unei analize a documentului se construiesc un dictionar al cuvintelor din acel document, apoi un dictionar cu toate cuvintele din toate documentele (index), fiecare cuvint avand referinte la toate documentele in care apare. La o cautare, termenii din interogare sunt cautati in dictionarul de cuvinte, identificandu-se documentele unde acestia apar. Documentele gasite sunt sortate in functie de relevanta si afisate. Exemple de seturi de date pe care pot actiona motoarele de cautare sunt documentele unei intreprinderi, arhiva unui ziar, o colectie de date personale sau chiar toate paginile de pe internet.

In timp ce la prima vedere design-ul unui motor de cautare pare relativ simplu, algoritmi de analiza, relevanta, sortare, optimizare etc sunt bazati pe tehnici avansate din programare, matematica si IT in general dupa cum putem observa si in aceste exemple de [algoritmi Microsoft](#) .

La nivel global, succesul acestor algoritmi se masoara in multe miliarde de dolari iar resursele hardware implicate pot ajunge de asemenea la dimensiuni considerabile, in functie de setul de date sub incidenta. Un exemplu este ferma de servere a Google, care cuprinde mai mult de [500.000 de servere](#), dupa ultimile estimari chiar circa 800.000.

#### 4. Istoric

Elemente de *teoria informatiei* si *data mining* se regasesc inca din anii '50 dar adevarata preocupare pentru motoarele de cautare a inceput odata cu aparitia web-ului la inceputul anilor '90. Primele programe de regasire a informatiei isi indeplineau partial rolul, avand in vedere ca seturile de date la vremea respectiva erau incomparabil mai mici fata de cele de astazi. Acestea au fost urmate curand de motoare de cautare semantice, dedicate web-ului, unele disparute astazi, cum ar fi Magellan, Excite, Infoseek, Inktomi, AltaVista sau Yahoo.

In anul 2000, doi studenti ai facultatii de matematica ai universitatii americane Stanford, avand o abordare exacta si comprehensiva, pun bazele [motorului de cautare Google](#). Acesta s-a dovedit imediat superior celor existente pe piata, in principal datorita relevantei si performantelor sale in termeni de viteza. De atunci acesta si-a continuat evolutia ascendenta ajungand astazi compania care practic domina internetul.

Este de mentionat evolutia la vremea respectiva a celorlalti doi mari jucatori pe piata actuala a motoarelor de cautare. In anul 2003, Yahoo realizeaza diferenta enorma de relevanta fata de motorul Google si inlocuieste motorul propriu cu cel al Google pe portalul sau (care la vremea respectiva era cea mai accesata pagina din lume) toate cautarile bazandu-se pe algoritmii acestuia din urma. Realizand ca pot domina foarte usor piata singuri, datorita produsului lor practic fara concurenta, in 2004 cei de la Google nu mai accepta aranjamentul cu Yahoo iar acestia revin la motorul propriu pe care continua sa il dezvolte si in prezent.

In tot acest timp Microsoft pierde teren si nu realizeaza importanta si potentialul din spatele motoarelor de cautare. Chiar daca portalul MSN avea motorul propriu inca din 1998, divizia de "Text mining & Search" a Microsoft este relativ mica si nu atentioneaza asupra cresterii vertiginoase a Google. Fiind deja in urma Yahoo ca cota de piata, ei ajung in acest moment pe locul trei, cu o cifra in jurul a 7% din cautarile de pe internet. Trezirea brusca in 2005 ii face sa accelereze lucrurile intern, incercand un rebrand al motorului propriu de cautare sub denumirea de *Live Search*. Relevanta acestuia este la fel de slaba ca inainte, succesul in termeni de cota de piata fiind practic nul.

In acest moment Microsoft intra in panica si incepe sa cumpere orice firma de pe piata care ofera algoritmi de cautare, pentru sume considerabile. In 2009 incearca iarasi o ofensiva si lanseaza *Bing*, avand un succes partial datorita relevantei un pic crescute insa inca departe de cea a Google, nici macar ca cea a Yahoo.

Lucrurile se complica si mai mult, Steve Balmer preia conducerea dupa plecarea lui Bill Gates si decide sa cumpere Yahoo pentru a "castiga" internetul prin forta bruta. Suma oferita la inceputul crizei este de 45 de miliarde de dolari, la scurt timp dupa aceea actiunile se prabusesc, CEO Yahoo isi da demisia iar [noua oferta](#) este considerabil mai mica, in jurul sumei de 17 miliarde de dolari.

Printre termenii contractului se regasesc specificatii exacte cu privire la motorul de cautare, sursele acestuia urmand sa fie puse la dispozitia Microsoft. In timpul scurs de dupa anulara aranjamentului cu Google, Yahoo si-a perfectionat algoritmi si invatand de la motorul de cautare pe care l-au folosit timp de un an, au reusit sa-si creasca relevanta proprie, ajungand aproximativ la 60% din cea a Google. In acest fel, prin [preluarea algoritmilor Yahoo](#), Microsoft spera sa castige teren in incercarea de a ajunge din urma Google. Cota de piata este in acest moment aproximativ 70% Google, 18% Yahoo, 12% Bing, in multe tari precum si in Romania Google avand peste 95%.

Exista aici un capitol intreg referitor la dezvoltarea regionala a motoarelor de cautare. In timp ce Google ofera domeniul "google.ro" si cautari bazate pe semantica limbii romane de cativa ani, Yahoo va implementa acest sistem abia catre sfarsitul anului 2010, iar Microsoft nu a anuntat inca o asemenea initiativa. Datorita alfabetului particular si grafiei, Google nu a putut penetra rapid in tari precum Rusia sau China, care au apucat sa-si dezvolte propriile motoare de cautare, Yandex respectiv Baidu. Tentativele de a intra pe piata celor doua tari sunt foarte agresive in prezent, ducand de exemplu la numeroase conflicte cu guvernul chinez sau la limitari legale impuse de guvernul de la Moscova. Bazandu-se pe relevanta lor superioara, este posibil ca pe viitor, odata intrati pe piata sa o si castige.

Au mai fost de asemenea incercari fara succes ale Germaniei si Frantei de a crea un motor de cautare european, cunoscut sub numele de Quaero (initial proiectul Theseus) dar de aceste proiecte incepute prin 2005-2006 nu s-a mai auzit nimic, fiind cel mai probabil esecuri. Circula zvonuri si despre un [motor de cautare global pregatit de guvernul rus](#), avand in spate resurse financiare puternice, rusii neimpacandu-se cu dominatia globala a Google.

Google si Microsoft au creat de asemenea [motoare de cautare locale](#), fara plata, care pot fi descarcate de pe internet si folosite, avand si o interfata de programare. (Aceste tooluri nu includ bineinteles algoritmi de cautare folositi in motorul de web, datorita protectiei impotriva piratariei si ingineriei inverse (reverse engineering). Sunt create insa pentru a descuraja aparitia altor motoare de cautare, care ar putea face ulterior concurenta.

## 5. Istoric romanesc

Anticipand evolutia pietei, Google a pus la dispozitia utilizatorilor un set de instrumente prin care aparent oricine isi poate crea un motor de cautare. De fapt este un fel de agregator al "engine"-ului Google pentru a afisa seturi de rezultate, fara a putea controla publicitatea dar permitandu-se redenumirea acestuia. Au aparut in acest fel un numar de pseudo-motoare de cautare, care practic ofera cautarea Google sub un alt nume. Acestea nu pot indexa seturi particulare de date atat timp cat ele nu sunt indexate deja de motorul Google. In acest fel, sub pretextul unui instrument util si usor de folosit, se descurajeaza crearea unor algoritmi particulari care ar putea ameninta vreodata pozitia Google pe piata.

Exista deja cateva asemenea [website-uri romanesti](#) care ofera asemenea cautari dar care nu sunt luate in considerare de vizitatori. In afara acestora, au mai fost cateva tentative dar foarte departe de a oferi relevanta ridicata sau o abordare profesionista, cele mai multe discutii de pe forumul de mai sus fiind relevante.

In afara de acestea, oricine poate restrange cautarea Google la un singur website prin asa numitul *Google Site Search*, astfel incat rezultatele sunt limitate la paginile din site. In acest fel se face inutila utilizarea altor motoare de cautare pentru un website, avand in vedere ca Site-Search-ul furnizeaza rezultate relevante.

In anul 2006, Adrian Sirbu patronul MediaPro, ii cere lui Orlando Nicoara, directorul diviziei de internet, crearea unui motor de cautare romanesc. Proiectul demareaza si dupa doi ani, beneficiind de resurse destul de consistente, 6-7 programatori alocati, largime de banda, servere, logistica etc, se lanseaza [Okidoki](#).

[Lansarea](#) si primirea acestuia pe piata au fost un esec, in primul rand relevanta acestuia fiind extrem de mica iar in al doilea rand vizitatorii romani nefiind dispusi sa renunte la motorul favorit, Google. O a doua tentativa de lansare, jumatate de an mai tarziu, a avut acelasi deznodamant. In plus, siteurile trustului precum "stirileprotv.ro" sau "protv.ro" etc sunt vizibil favorizate de okidoki.ro, fapt pe care creatorii Google l-au evitat de la bun inceput, ramanand fideli neutralitatii.

Intre timp Okidoki este prezent pe toate site-urile Mediapro cu toate ca aceeași cautare se putea realiza și cu Google Site Search. Acesta este și motivul pentru care nu s-au extins la alte website-uri în afara trustului, însă probabil vor face și asta mai devreme sau mai târziu.

Proiectul continua și nu cu mult timp în urmă s-a realizat un [contract cu Romtelecom](#) pentru personalizarea paginii de eroare DNS cu ajutorul Okidoki, percepută însă negativ de către utilizatori. În prezent probabil că Okidoki este foarte puțin folosit de publicul larg, relevanța lui fiind la nivelul motoarelor de cautare ale caror surse se găsesc pe internet, însă personalizarea acestuia pe limba română, prezentă pe siteurile Mediapro și resursele trustului din spate îi pot asigura un oarecare viitor.

## 6. Potentialul motoarelor de cautare

Sub simplitatea rezultatele cautarilor pe internet se ascunde un potențial enorm, cel care controlează motorul de cautare major controlând practic internetul. Diferența între a fi pe poziția 10 (prima pagină) și 11 (a doua pagină) la o cautare relevantă pentru un domeniu se poate măsura în milioane de dolari, foarte multe firme investind masiv într-o bună imagine și expunere pe web prin folosirea instrumentelor puse la dispoziție de Google precum *AdWords* și *AdSense*. Acestea sunt *publicitate contextuală* în funcție de cuvintele cheie folosite și intervin odată cu apariția rezultatelor și sunt legate direct de acestea, în acest fel publicitatea fiind intrinsec legată de motorul de cautare. Plătești să fii în rezultate, vezi rezultatele plătite la cautare sau la accesarea siteurilor.

Traficul oricărui website este direct afectat de poziția acestuia în motorul de cautare cel mai des accesat. O mare parte din vizitatorii unei publicații media ajung prin intermediul motoarelor de cautare, aceștia optimizându-și paginile pentru o mai bună expunere. Vizitatorii au ajuns să tiparească în Google numele site-ului pe care doresc să ajungă, în loc să îl acceseze direct. Multe firme care se bazează exclusiv pe internet au dat faliment când Google și-a restructurat algoritmi și a schimbat poziția rezultatelor. Mai mult, Google înregistrează cautarile efectuate și o multitudine de date despre utilizatori, prin referințe încrucisate cu alte website-uri precum *youtube.com* sau *gmail.com* al cărui posesor este, cunoscând în acest mod tendințele și necesitățile înaintea tuturor, făcând profilul fiecărui vizitator și având o imagine exactă a internetului.

Este suficient să ne uităm la statisticile puse la dispoziție prin instrumente precum Google Analytics sau chiar statisticile YouTube, pentru a realiza că sunt suficiente informații pentru a face profilul complet al unui utilizator de internet, incluzând datele personale, pozele sale, ce a căutat în ultimele luni, ce preferințe și preocupări are, ce filme a văzut, ce site-uri a vizitat, ce comentarii a postat, ce prieteni are și poate chiar și ce corespondență a avut. Să nu uităm că majoritatea site-urilor își înglobează în paginile sale scripturile "*Analytics*", acestea executându-se pe serverele Google, codul HTML fiind întors de aceștia.

Este greu de spus cum va arăta viitorul, însă odată cu creșterea rapidă a informației și crearea continuă de conținut de către vizitatori, prin intermediul blogurilor, comentariilor, paginilor personale etc filtrarea acestor date [va fi tot mai dificilă](#) și probabil se va apela tot mai des la factorul uman.

În afara de acestea este posibil să ne întâlnim cu o *specializare a motoarelor de cautare*, astfel încât vizitatorii probabil vor migra către motoarele de cautare locale sau subnivele de cautare pe anumite domenii. De exemplu ca utilizator, în loc să caut pe Google cazare într-o anumită stațiune, unde se vor întoarce haotic prea multe rezultate nerelevante, s-ar putea să prefer să caut pe un portal de turism același lucru, unde pot ordona după data sau relevanță. Sau s-ar putea să vreau să caut rezultate numai în paginile ziarelor românești, ignorând blogurile sau alte pagini neinteresante sau realizate de amatori.

Începând cu anul 2008, în urma lansării motorului de cautare *Cuil* care pretindea că are cele mai multe pagini indexate, Google și-a revizuit algoritmul și a negociat o parte din relevanță pentru un volum mai mare de pagini indexate. Așa cum s-a văzut, tentativa de a indexa toate paginile din internet are un cost destul de mare, fermele de servere ajungând la sute de mii de calculatoare.

In incercarea Google de a-si mentine monopolul, sau in cea a Microsoft de a-l dobandi, cele doua companii vor cumpara fara ezitare firme mici care au ceva de oferit sau vin cu o imbunatatire sau o inovatie. Chiar si o cota de piata locala de 1% ar fi suficient pentru a trezi interesul acestora, problema este ca cei 1% sunt extrem de greu de realizat mai ales daca nu ai ceva de oferit, si in special **relevanta** - unitatea de masura a calitatii unui motor de cautare. Informatie de filtrat va exista mereu, si mai mult decat atat, aceasta se acumuleaza continuu, generand noi probleme si necesitand noi solutii.

## 7. Premizele Zamolxe

Fiind conectati permanent la mediul online si beneficiind de pregatirea necesara, am inceput sa analizam motoarele de cautare si posibilitatea de obtine o pozitie cat mai buna in acestea. Am realizat ca exista anumiti parametrii care fac ca un motor sa devina relevant si niste principii matematice pe care trebuie sa te bazezi.

Cand ne-am dat seama de elementele care pot conduce la un motor de cautare relevant, am demarat constructia Zamolxe, la realizarea prototipului avand satisfactia sa descoperim ca nu ne-am inselat, obtinand la cautari o relevanta similara cu cea a Google pe seturi restranse de date.

Dezvoltarea algoritmilor s-a intins de-a lungul mai multor ani, fiind bazata pe numeroase observatii, calcule si teste.

Motivatia care ne impinge sa speram in potentialul Zamolxe consta in soliditatea algoritmilor acestuia dovedita prin teste precum si in faptul ca internetul romanesc este in plina ascensiune volumica. Chiar daca lucrurile s-au asezat aparent din punct de vedere al influentei, experienta arata ca schimbari majore se vor produce oricum. Suntem convinsi de posibilitatea realizarii unui motor de cautare romanesc cu relevanta ridicata si sunt de parere ca un brand, o logistica si o tehnologie create acum in acest domeniu ar putea avea un impact substantial in 5-10 ani. Explozia internetului a inceput practic in jurul anului 2000, motoarele de cautare mari incep deja sa ignore informatia mai veche de 4-5 ani si relevanta scade vizibil de la an la an. Este de anticipat ca odata cu acumularea informatiei sa se creeze o nisa pentru motoare de cautare locale si specializate, poate si pentru Zamolxe.

## 8. Zamolxe – date tehnice

Tehnologia Zamolxe este integral dezvoltata on-site si nu foloseste alte librarii externe de cautare utilizand doar algoritmi de indexare si cautare proprii. Regasirea informatiei este realizata tinand cont de complexitatea si particularitatile limbii romane, motorul avind urmatoarele caracteristici:

- relevanta inalta - motorul "gandeste", intuind cerinta utilizatorului;
- viteza si optimizare - informatia este regasita instantaneu in volum mare de date;
- cautare semantica full-text;
- analiza frazelor si a termenilor in limba romana;
- diacritice romanesti si lematizarea formelor flexionare;
- baza de date interna de nume romanesti;
- algoritmi proprii de indexare si cautare;

Pentru consistenta documentului vom adauga cateva caracteristici tehnice, aceste date fiind insa predispuise la schimbare in timpul dezvoltarii.

Structura interna a Zamolxe este formata din 3 componente de baza, crawlerul, indexerul si searcher-ul. *Crawlerul* este componenta care parcurge si descarca paginile unui website, fiind proiectat pentru accesul cu viteza ridicata, in sistem multithreading, pornind de la pagina principala. Output-ul acestei componente rezida intr-o serie de fisiere compresate, continand paginile website-ului precum si o serie de informatii aditionale legate de acestea. In cazul in care se dispune de o latime de banda mai mare, crawlerul poate fi extins foarte usor la o capacitate mai mare, fie crescand numarul de threaduri (fire) de acces, fie distribuindu-l pe mai multe masini.

*Indexerul* cuprinde analizorul de pagini avand si o componenta aditionala denumita *Sorter*. Acesta sorteaza dictionarul obtinut in urma analizei paginilor pentru o viteza mai mare de acces a celei de-a treia componente, Searcher-ul. Viteza acestora este satisfacatoare dar poate fi crescuta prin folosirea unui hardware superior. De asemenea Sorterul nu este inca in forma finala, pentru un index de ordinul zecilor de Gigabytes trebuind sa fie implementat un algoritm diferit.

*Searcherul* este componenta cea mai importanta, fiind construita ca un compromis pentru viteza. Orice cod in plus inseamna mai multe milisecunde, implicit un timp de asteptare care face folosirea motorului neutilizabila in conditiile de astazi. Obtinerea unei relevante ridicate in aceste conditii se bazeaza pe algoritmii folositi. La un motor de cautare, mai mult cod sursa sau o masina mai mare nu inseamna neaparat ridicarea relevantei atata vreme cat algoritmul folosit nu este suficient de bun.

Pentru demonstratie se va folosi baza de date Adevarul cuprinzand 413.000 de pagini, timpul de acces la rezultate fiind de circa 30 de milisecunde. Indexul acestor pagini are o dimensiune totala de aproximativ 1GB, memoria in timpul cautarilor atingand 600 de Megabytes.

Pentru o buna observare si obiectivitate se vor compara rezultatele obtinute pentru aceleasi interogari aplicate acelorasi seturi de date de catre diverse motoare de cautare. Intrucat Zamolxe a indexat doar "adevarul.ro", vom aplica interogari aleatoare pentru acest website urmatoarelor motoare de cautare: Google, Yahoo, Bing, Cuil, Okidoki. Testele preliminare arata ca relevanta Zamolxe este comparabila cu Google sau superioara, net superioara celorlalte motoare de cautare din esantionul de test.

## 9. Posibilitati si necesitati

Exista nenumarate domenii unde optimizarea regasirii informatiei poate fi aplicata. Din pacate nu prea exista oferta pentru optimizari de cautare iar pe de alta parte companiile nu realizeaza cata informatie se pierde prin folosirea unui motor de cautare nerelevant.

O companie cu un volum mare de documente prefera sa cumpere un manager de documente care are deja inglobat un sistem de cautare exacta. In cazul a 100 de rezultate apropiate iti permiti sa petreci un timp analizand fiecare document sau titlu pana ajungi la cel dorit, insa daca sunt 1000 de documente apropiate de ceea ce cauti, lucrurile devin mai dificile. In plus, cautarea exacta poate rata rezultate, documentul dorit putand sa contina termeni din familia lexicala a interogarii si nu termenii exacti din interogare. In cazul folosirii unui motor ca Zamolxe, documentul dorit este intors cel mai probabil ca prim rezultat. Un ziarist care se documenteaza pentru un articol poate gasi instantaneu toate articolele din ultimii 10 ani referitoare la un subiect cautand in arhiva ziarului, o lege poate fi gasita instantaneu cautand termenii pe care ii contine in toate legile in ultima 100 de ani. Acestea sunt doar exemple, cautari similare se desfasoara in prezent insa nici macar nu se stie cate rezultate nu apar datorita cautarii exacte.

Un exemplu ar fi arhiva ziarului Adevarul unde cautarea exacta intoarce rezultate total nerelevante (ex "programul de guvernare"). Mai potrivita ar fi fost utilizarea Google Site Search, Adevarul avand

250.000 de pagini in indexul Google. O indexare personalizata, de exemplu cu Zamolxe, ar efectua cautarea pe mai mult de 400.000 de pagini din arhiva acestui ziar, cu o relevanta ridicata.

Este dificil de spus ce succes ar putea avea un alt motor de cautare pe piata, experienta Okidoki nu este incurajatoare dar Okidoki nici nu a avut foarte multe de oferit, relevanta lor ridicandu-se si in prezent la nivelul unui motor de cautare ale carui coduri sursa se gasesc liber pe internet. Abordarea initiala de a "incerca sa bata Google in Romania", declaratiile in acest sens au constituit o strategie gresita si chiar un nerealism de situatie care i-au facut sa arate penibil de la bun inceput, alungand vizitatorii. Lansarea prematura, in stadiul beta, cand relevanta era practic inexistentă a atras o gramada de comentarii negative si trag inca consecintele rezultate din inertia acestei prime pareri. Poate ar fi mult mai util sa indexeze unul cate unul fiecare ziar din Romania, sa-si dovedeasca utilitatea, sa intre in atentia unui nume mare precum Google sau Microsoft. Momentan dezvoltarea Okidoki continua, in timp ce ei cauta noi pietele si solutii pentru a se impune.

Dezvoltarea unui motor de cautare se poate concentra pe algoritmi sau pe ceva inovator care ar putea fi vandut mai tarziu. Nenumaratele probleme datorate diacriticelor romanesti, rezolvate partial de Zamolxe, ar putea de asemenea constitui un atu al acestuia, alaturi de relevanta inalta. O combinatie intre un motor de cautare romanesc orientat catre media si un engine local gen [dtsearch.com](http://dtsearch.com) ar fi de preferat ca abordare initiala.

Resursele necesare pentru a putea crea aceste premize nu sunt insa la indemana unei firme mici si nici chiar beneficiind de ele nu exista o garantie a succesului. Numele mari de pe piata sint in expansiune continua, razboiul intre ele in aria de cautari este in plina desfasurare, ele dorind si sa acapareze pietele locale. Zamolxe are o parte din ceea ce lor le lipseste si poate avea si mai mult. Crearea unui brand "Zamolxe" ar fi de asemenea posibila, cu ajutorul domeniilor "zamolxe.ro" si "zamolxe.eu".

## 10. Propunere de colaborare

Dezvoltarea Zamolxe ar putea continua individual, ca si pana acum, va fi insa nevoie de multa sansa pentru a obtine expunere pe piata romaneasca, neexistand resurse pentru a impune acest nume si nici pentru a dezvolta alte module intr-un timp util. Ar exista posibilitatea de a-l oferi spre utilizare unor institutii, publicatii media, website-uri etc si in cel mai fericit caz ar atrage atentia si poate s-ar putea vinde pentru o suma modica. Dezvoltarea lui ar continua in paralel cu alte proiecte, ceea ce ar lungi si mai mult perioada de productie, pana la a-l face necompetitiv.

De preferat este asocierea cu o institutie cu forta financiara si preocupari in domeniul internetului, care intelege amploarea si perspectivele unui asemenea proiect. Intr-o perioada relativ scurta s-ar constata primele rezultate si ar putea crea premisele unor venituri pe termen lung. Fie prin vanzari directe ale publicitatii sau rezultatelor, fie prin cumpararea de catre o companie interesata de progresele facute in domeniul relevantei sau cautarilor locale romanesti (Microsoft, Yahoo), fie prin utilizarea rezultatelor indexarilor.

In timp ce resursele pentru o asemenea investitie nu sunt greu de estimat, perspectivele si oportunitatile care pot aparea sunt prematur de anticipat ele depinzand de progresele facute, de dinamica pietei romanesti, eficienta marketingului etc. Este probabil ca in primul an, in timpul construirii brandului si al dezvoltarii, sa nu existe nici un fel de profit in afara utilizarii proprii si al consolidarii prezentei online. In cazul unui succes insa, acesta poate avea valori mai mult decat substantiale.