# To exploit the collective intelligence thanks to the Co-operative replication

Luigi Lancieri, Nicolas Berthier Bonnel, Ludovic Stumme
France Telecom R&D
Rue des coutures 14000 Caen FRANCE
Luigi.lancieri@francetelecom.com
Ludovic.stumme@monetique.ensicaen.ismra.fr
Nicolas.berthier-bonnel@monetique.ensicaen.ismra.fr

## Abstract

*For some time we work on the interest of information recycling obtained by implicit co-operation and in particular on the recycling of the proxy caches contents. The first phase of this work led to a system, which indexed the contents of HTML pages got from a proxy cache. The objective was to provide a search engine implicitly centred on an intranet users' fields of interests. The promising results of this work briefly described below pushed us to further investigation and in particular to recycle the reusable heavy size objects contained in the cache. We shows in this paper that the strategy of reusing already downloaded information provide interesting advantages at low cost. In particular to speed up average accesses to Internet, to improve information search and to contribute to save bandwidth on Internet. Moreover this approach comes with in the conceptual frameworks of informational ecology and collective intelligence.*

**Keywords**: Internet, Proxy caches, co-operation, Reusability, QoS (Quality of service), Search engine

## 1. Introductions

The collective intelligence is a well-known phenomenon in the living world and in particular in the animals or humans societies. The basic idea is that the co-operation between the individuals is more productive than the sum of the individual actions. Internet is a place where the informational co-operation and interactions are omnipresent, which makes an ideal ground of experimentation of it. A certain number of services as the news groups for examples have an operation mode based on the more or less implicit co-operation. In the same way, Web sites cross references connecting the ones to the others information of different sources make of Internet the result of a huge co-operation. From this point of view we can compare the network with a big informational ecosystem [1] [2]. Our objective is to understand and to reuse this co-operation. As we will see below, the proxy-cache is an interesting component because its content is the result of the users' implicit co-operation and that this content is very easy to reuse. The guiding principle of the system described in this paper is the re-use of the heavy size contents stored in the proxy caches. The objective is to increase the data accesses speed from Internet and the selectivity of this access (make easy to find what you are searching for). These two objectives contribute to a more effective access to information on the Web. To replicate information on servers close to the users is a good solution to accelerate accesses. The entire problem is of knowing what information is useful to replicate and how to feed these servers so that the economic equilibrium is favourable. In order to solve this problem we propose a method, which exploits the implicit co-operation of the proxy-caches users' to feed http servers located near them.

## 2. Description of the proposed system

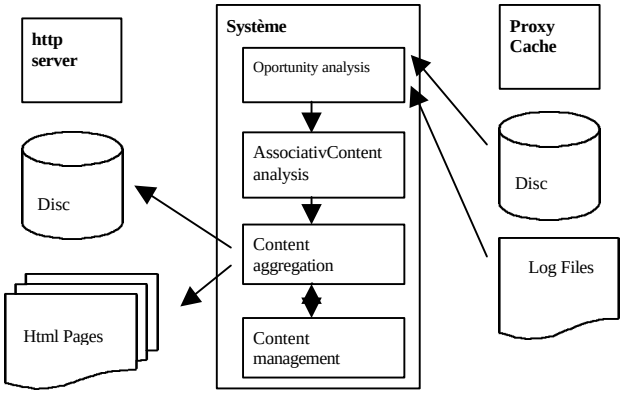


**Figure 1** : Functional Architecture

The user sees the suggested system as a traditional HTTP server that he reaches via a regular Web browser. The HTML pages located on this server are automatically generated and contain links towards heavy size objects and its descriptions recycled from the cache. At the beginning the cache contents (objects, associated HTML pages and log files) are analysed to select useful objects. We keep only easily reusable content (mp3, mpg, Doc., avi, jpg, etc) dropping more complex one (dll, vbx, dowry, etc). Contrary to a regular cache policy, we consider that the cost of storage is less heavy to support than the management of the replicates. It is thus not interesting to recycle small size objects since they can be downloaded rather quickly from the origin server. When the useful

objects have been identified, it is necessary to recopy them and also look for some descriptive key words. This can be carried out in several ways (URL key words, analyses of Web page containing the link to the useful object, extraction of the informative fields in MP3 files, etc). We will not go into the details here but the extraction of the multimedia objects description is not an easy problem. Let says also that before to copy an object from the cache it is necessary to do some processing on it like to remove the special heading generated by the cache. One of the important differences between our server and a traditional one it is the automatic management of the objects time to leave. This is necessary because the objects feed process is continuous. The server takes into account the characteristics of the object (size, category, thematic content, etc.) and the interactions from the users (a number of access, variation compared to the topic of interests, etc). (See [3] for more details on these aspects.) With all these elements the server compute a priority function for each object which will make it possible to decide which one to keep (multicriterion LRU stack). One of the guiding principles of our system is that the disc size is very large (the target is several hundreds of G Bytes), which implies that the objects lifespan is at least 1 to 2 weeks long. It is an important difference compared to a regular cache (long retention time). Let says also that it is possible to use such a system over a co-operative cache network in order to extend the advantages to a larger users' community.

# 3. Results

We present here some results of the use of the described system. The preliminary experiment was undertaken during 1 month with two 3 G Byte caches (SQUID) on a 250 peoples' Intranet. The selected content of these 2 caches is recopied on a 4 Go http server. This experiment involved 15 volunteer users of the server. How we said it, the combination of the http sever and the cache increases the hit rate (reuse of local objects) and mechanically the access speed to information. The Hit rate becomes here a combination of the implicit action (access to the cache) of the users and explicit one (access to the server). The second aspect of the advantages of this system is the fact that the contents are implicitly oriented towards the users' interest topics what makes it very interesting to reuse (search time reduction, co-operative filtering, proposals to users, etc). Each member of the team can benefit from the efforts of the others, from where the idea of collective intelligence. We develop these two aspects by taking into account several experiments related to the same concept that proceeded over several years.

## Influence on the access speed to information

The traffic logs analysis of the French national cache Renater [4] shows that the access to a local server is on average 300 times faster than the access to a distant one[1]. In addition, these logs show that the latency time tends to increase (approximately +70 % over the year 2000). The analysis which follows aim to show that the combined action of the cache and the server significantly increase the objects re-use rate and consequently the average access speed to Internet. We start by analysing the behaviour of the cache, then that of the server and finally the combined action of both.

The table that follows gives us a synthesis of the cache activity over the considered period of time with the selected mimes categories. It is noted that potentially the reusable quantity of data is relatively high. Indeed the users carried out requests equivalent to 11 GB of data where 79 % are unique requests. It is also noted that the hit rate on the big size files is relatively weak 14 % on average against 36 % on overall files. This implies that the heavy objects are re-used very little in a standard cache. It is a pity because they are the more difficult and longest to obtain. One of the reasons possibly explaining this fact is that the heavy objects are proportionally much less frequent than the small one (e.g 93 % of the images jpg make less than 40 KB - ergonomic size of a picture - with an average of 11 KB). [2]

| type | Req miss | Size miss | Req hit | Size hit | Req total | Size total | Hit Req % | Hit Size % | # uniq |
|---|---|---|---|---|---|---|---|---|---|
| mp3 | 421 | 651 | 88 | 98 | 555 | 757 | 16 | 13,0 | 461 |
| jpg>40 | 8 895 | 722 | 1 503 | 126 | 10 624 | 857 | 14 | 14,6 | 9 391 |
| mpg | 3 850 | 4 026 | 243 | 302 | 4 221 | 4 365 | 6 | 6,9 | 3 766 |
| msvideo | 187 | 808 | 13 | 66 | 213 | 880 | 6 | 7,5 | 178 |
| zip | 946 | 2 377 | 97 | 77 | 1 051 | 2 459 | 9 | 3,1 | 997 |
| qTime | 182 | 484 | 85 | 410 | 303 | 895 | 28 | 45,8 | 149 |
| pdf | 2 355 | 498 | 842 | 182 | 4 003 | 697 | 21 | 26,1 | 1 583 |
| **total** | **16 836** | **9 567** | **2 871** | **1 262** | **20 970** | **10 910** | **14** | **11,6** | **16 525** |

**Table 1**: One-month caches activity, on the standard major size mime categories, Size in Mb

The other reason is that these objects are badly indexed by the search engines. So that, in some case its discovery is almost done by chance.

[1] One year average access on global Internet from France. The average fetch time of a cached object first byte vary from 9 to 25 ms in a local cache Intranet against from 1 to 7 sec when the cache fetch it from the origin server.
[2] This is because the function linking object size and access popularity is a power law (Zipf Law).

Table 3 gives indications on the recycled objects and their level of re-use on the server by the 15 volunteers. It is noted that the utilisation ratio of information is higher than that of the cache (480 requests per user over the period against 83 for the cache). This is not surprising taking into account the remark made previously. Indeed, our server simplifies the access to the heavy objects since it presents them explicitly. Furthermore the download is extremely fast so the user less hesitates to get these objects because it does not have to wait long minutes for contents which have perhaps low worth.

| type | # uniq | Size | averag Size | # total Used | Size |
|---|---|---|---|---|---|
| mp3 | 242 | 526 | 2,1754 | 323 | 1070 |
| jpg>40 | 5779 | 473 | 0,0819 | 3997 | 327 |
| mpg | 869 | 890 | 1,0244 | 2295 | 3877 |
| msvideo | 58 | 150 | 2,5777 | 174 | 483 |
| zip | 477 | 916 | 1,9212 | 106 | 337 |
| qTime | 69 | 144 | 2,0941 | 80 | 225 |
| pdf | 669 | 352 | 0,5268 | 225 | 165 |
| **total** | **8163** | **3453** | **10** | **7200** | **6484** |

**Table 2:** Uses of recycled content on the server

| type | serv Unique | Cache Uniqu | Cachable % |
|---|---|---|---|
| mp3 | 242 | 461 | 52,494577 |
| jpg>40 | 5779 | 9391 | 61,5376424 |
| mpg | 869 | 3766 | 23,0748805 |
| msvideo | 58 | 178 | 32,5842697 |
| zip | 477 | 997 | 47,8435306 |
| qTime | 69 | 149 | 46,3087248 |
| pdf | 669 | 1583 | 42,2615287 |
| **total** | **8163** | **16525** | **49,397882** |

**Table 3**:Cachahbility of big size objects

One can compute the cachability rate because the objects stored on the server are all cacheable. Indeed, the recopy process intervenes immediately after the object is in the cache, which implies that the cache elimination mechanism does not have time to act. This is important because a very variable part of Web objects are dynamic one and thus not cacheable. We will see below that this cachability is also fundamental from the point of view of the intellectual property. One finds the cachability rate by computing the ratio of the single objects on the server divided by the single objects in Access log of the caches (users' requests). One finds a value of about 50 % that are lower than that of the html pages which is about 90 % (that is also near the average value of overall objects, see [5] for a complete study on cachability). The following table shows the increase in the hit rate perceived here as the increase in the re-use of the objects due to the cache / server association, from where increase in the data access speed.

| Cache 250 users | | | | | Server 15 users | | Global | |
|---|---|---|---|---|---|---|---|---|
| type | # Hit | # Total | Raw Hit % | Reuse % | # Unique | # Total | Raw Hit % | Reuse % |
| mp3 | 88 | 555 | 16 | 30 | 242 | 323 | 46,8 | 66,9 |
| jpg>40 | 1503 | 10624 | 14 | 23 | 5779 | 3997 | 37,6 | 52,2 |
| mpg | 243 | 4221 | 6 | 25 | 869 | 2295 | 39,0 | 77,6 |
| msvideo | 13 | 213 | 6 | 19 | 58 | 174 | 48,3 | 76,8 |
| zip | 97 | 1051 | 9 | 19 | 477 | 106 | 17,5 | 33,3 |
| qTime | 85 | 303 | 28 | 61 | 69 | 80 | 43,1 | 74,9 |
| pdf | 842 | 4003 | 21 | 50 | 669 | 225 | 25,2 | 55,7 |
| **total** | **2871** | **20970** | **13,7** | **27,7** | **8163** | **7200** | **35,8** | **57,4** |

The Raw Hit rate is computed by dividing the cache hit requests added to the server requests (that are in fact hit on cached objects) by the total accesses (total cache plus total server). The reuses rate is computed by the same formula except that the « total cache » is reduced to the cacheable part.

**Table 4**: synthetic analysis of the total system

## Influence on the information access selectivity

One of the strong points of this system is its adaptability. Indeed, the contents of the cache will move and follow the users. So that, it constitutes a reduced information database centred on the users' mains interests topics. The first version of this system was only able to index html pages (as a search engine but restricted to the cache content) but it shows interesting performances. [6][3]. We had asked 7 users to make ambiguous requests (e.g. Network) and to give an evaluation value reflecting the level of interest for the first 20 results provided compared to HotBot Lycos. The average values are 3.2/5 against 0.7/5 for Hot Bot. The size of the files corresponding to URLs returned by our system were big size textual files compared with those of Hot Bot (128,2 KB [3]on average against 13,4 KB, 10 times more). The answers are centered on the network Internet techniques which are the field of competence of our company (55 % of the answers are 1.7 Mo on the whole). In comparison, the firsts answers of Hot Bot are relatively heterogeneous and strongly trade and advertisement directed with very short contents (65 % of the answers are 97 KB on the whole). The first relevant response provided by Hot Bot is to the 12 Th position whereas on the first position our system provides documents with strong technical contents. One will note also the difference of the number of provided results, 300 URLs whereas Hot Bot in provided 50 000. Actually, these results are not surprising because the contents of the cache follow the mains users' interests. The negative point is that the implicit filtering used by this system is only effective when research relates to an interest topic of the group, outside of this area the results are extremely weak.

[3] A 128 kb file represente around 50 A4 textual pages.

# 4. Example of implicit co-operation exploitation on Internet

Collective Intelligence was the subject of lots of interesting works over the last decade. Briefly we can split these studies in 3 mains areas the first one is strongly social an human science oriented [26], the second one is strongly artificial system oriented [27] and the last one is a more equilibrate mixture of the two first one and basically involves Human-Systems interactions [3]. This is mainly the focus of our studies. The implicit co-operation is very present on Internet. It intervenes for example in the data routing, in Web contents generation or replication. We will insist here on the last 2 aspects but further details on the generalisation of implicit co-operation on Internet are approached in [3].

The cache is the simplest component based on the implicit co-operation principles and the information re-use. On the first access to a document, this one is memorised and returned when a second user ask it again from where saving of data transmission time and network bandwidth. For a question of optimisation the caches store only the contents that are not too heavy moreover, generally the average storage period is still few days except if the object is very often asked. The principal consequence in relation to our concerns is that it is difficult to foresee the documents, which will be, stored so as their lifespan, because they depend on the users and Web informational interactions (interests topics, numbers of access per days, etc). Contrary to the caches, the contents of the Web servers and the mirrors are completely deterministic. I.e. the administrator must take the initiative and explicitly store information and consequently he controls all the parameters related to the contents of these servers (lifespan, quantity of replicates, localisation, etc). Moreover, the mirror sites management is often mechanical (in some case manual,) i.e. that it is systematic, and generally implies identical organisations of the contents between the originals server and the mirrors. The CDN (Content Delivery Network) is an improved version of the mirrors of which it solves some gaps. It constitutes at the base. The system of CDN is a distributed architecture of storage components (mirrors or caches) transparent for the user and makes only replicate the organisation and the contents of previously selected origin servers (contents determined in advance). Its autonomy when it exists locates more in the management of the replication that in the constitution of the contents. In addition to these approaches now traditional, the systems of replications gave places to much academic work whose objective was to exploit the implicit co-operation. P. Rochat et all [6] propose to take into account the cultural influence of the users to optimise the performances of the caches. The basic idea is that there is a strong link between the words contained in the URL and the users' interest's topics. LSAM Project (Large Scale Active Middleware) [7] uses the multicast to diffuse Web pages on the level of the co-operating caches while taking into account community of interests. (see also [8] [9]).

| | Autonomy Constitution of the contents | Document Size | Facility of Contents management | Average lifespan | Choice Selectivity | Averg. QoS |
|---|---|---|---|---|---|---|
| Caches | + + + | + | - | + + | - | + + + |
| origin Servers | + | + + + | + + + | + + + | + | + |
| Mirrors | + | + + + | + + | + + + | + | + |
| CDN | + + | + + + | + + | + + | + | + + |
| Search Engines | + + + | - | + | - | + + + | - |
| System suggested | + + + | + + + | + + | + + + | + + + | + + + |

In these systems, the implicit co-operation is exploited to accelerate the accesses but seen of the end-user they are as transparent as caches. This table gives a comparative synthetic view of the majors replication systems in our concerns perspective.

**Table 5:** Comparative between our system and various traditional components.

Another way of taking part of implicit co-operation is that of the direct exploitation of the Web content and user activities. Project FET [10] for example, proposes to study and exploit the phenomena of collective intelligence on the Web on the same basis as that put in operation in the neural networks (associative links). Are taken into account the users' observed variables like thematic profiles, the reciprocity of connections, the time of consultations, etc). The analysis of this information can be used to make recommendations to users or to drive self-organisation of the network. K Nakata and all [11] used the principle of the co-operation to extract information related to cross referencing in shared databases. This approach mixes implicit and explicit co-operation. Each user marks key words important for him in the consulted documents. This database is then analysed by using A.I. techniques to produce an index of concept that will be used amongst other things to make easy the search of information. L.Terveen and all [12] also worked on the evaluation of collaboration in the Web. They analysed the cross-referencing between the Webs pages recovered concerning about thirty topics. They used it as a Web sites quality indicator. They consider that this co-referencing is an indication of emergent collaboration (density of the associated graph). They showed for example that co-referencing was very limited in a commercial context. F.Helyghen [13] explores in his paper the phenomena of collective intelligence and its computation in the Web. It proposes the construction of a collective mental Map containing the state of problems to solve, the possible actions and the preferred ones. It calculates in particular a co-occurrence matrix of the links contained in the Web pages consulted by the users, as well as the analysis of their browsing path. All this information is used to produce collaborative filters that propose new relevant Web sites to the users or to feed an optimised research agent.

Another way to exploit collective intelligence from the Web is to re-aggregate part of distributed Web pages on the basis of users' topics of interests or usage. Unfortunately, this individuality badly put up with the need for many contents suppliers' that need to reach a large audience and thus proposing contents sometimes very heterogeneous. Moreover the studies show that the majority of the users revisit few Web sites Web (around 5 to 10) compared with the totality of the visited sites (see state of the art in [3]). This was partially taken into account by the contents suppliers' who proposed personalised accesses (MyAltavista, MyYahoo, etc). The idea of aggregation, further goes, it involve the idea of collecting some parts of dispersed sites or Web pages and of associating them on a single access. The principal difficulty is to split Web pages in semantically homogeneous parts and to describe them well. These descriptions will make it possible to sort contents or to guide the process of re-assembly. These various stages can be more or less automated by systems based on the artificial intelligence also using XML standard. ([14], [15], [16], [17], [18], [19], [20], [21]).

## 5. Legal and ethical aspects

The collect and the reconditioning of objects downloaded from the Web pose several problems with respect to the users (e.g. to keep the confidentiality of the access) and to the owner or the supplier of the objects (intellectual property). These questions are important and without exhausting the subject here we can say that overall the answers strongly depend on the context. Indeed, the case of use in Intranet with a multi-site multinational corporation specific content is different from that of public ISP not controlled contents from Internet. First case is traditional and can be regulated by a company charter knowing that from the law point of view the problem is not more difficult than that of the traditional access to Internet. These networks are sometimes very large with such a quantity of information that one finds the same problems as on Internet even if the contents are controlled better. In addition, the majority of countries legislation allows the company to analyse the traces of connections if the employees are well informed. So, the implementation of this system makes it possible to make benefit from implicit collaboration the members of the company on "clean" contents. In the case of an open Internet access it should be pointed out that the contents of the system are not different from the one downloaded by the employees and present on their machine. This point of considering our system does not modify enormously the responsibility for the company that is already engaged by the contents already present on the machine of their employees. However, in some cases, the law forces the companies to make comply with precise rules that imply for example, the use of filtering systems to eliminate not allowed contents. The case of the general public I.S.P is more delicate. For reasons of ethics it is always essential that the users be well informed and that they validate the use that is made with their browsing traces. If necessary the confidentiality can be preserved thanks to the use of pseudo that makes the traces anonymous.

Another problem is that to satisfy the contents suppliers who are often remunerated by the users' hits, which imposes that the object is consulted on the origin server. The system described is not incompatible with this constraint. As we said it above, all the objects are not cached. In fact, the contents suppliers can avoid the cachability of an object. This can be done in particular in the HTTP response with the appropriate header [24]. So a content supplier can freely decide if its contents could or not be exploited on others servers. If the contents suppliers accept the re-use, the system allows to automatically sending back users' hits (with out object transfer over the network). This method is rather common (see [22][23]) in the replications systems world. In addition, the URL of origin server is shown to users with the copied object. These problems of property rights are important. The case Napster showed that information owners' have the responsibility to mark their contents and that Napster had the responsibility for filter the copyrighted contents. In the same philosophy we can implement filtering systems which make it possible to avoid certain type of contents. It is also possible to make money redistribution in a commercial partnership context. Thus, one can say that a certain number of guarantees are taken to satisfy the contents supplier as well as the final user.

## 6. Conclusion

Within sight of the results of this experiment it appears that this system is effective and with little cost. He combines user' explicit and implicit collaboration by exploiting existing components (cache) and a generic interface (e.g navigator, CGI, etc.). This approach that limits the specific developments makes it possible to reduce the costs of putting in operation and maintenance of the product. The results often integrate objects contained on recent sites, which are not indexed yet by the traditional search engines (case of the sites transmitted of mouth to ears). [4] Moreover by recycling already downloaded objects this system contribute to a necessary effort of reducing Internet bottleneck for the benefit of the majority. It remains despite everything, some delicate aspects like the problem of storage of illegal objects with the control is likely to be heavy in a completely open context. Moreover, the possibilities still limited in terms of

[4] One know that 10 % of the Web are indexed by search engine [25]

characterisation of multimedia objects limit the performance of the re-use even if that is much simpler than on open Internet and taking into account the fact that the recopied contents are fewer and users centred. In spite of these handicaps we think that our approach is a solution to deepen to fight against the growing slowness and the informational disorder of the Internet network.

## 7. Acknowledgement

Special thanks for Sebastien Chastang, Nicolas Durand, Samuel Legouix, Nicolas Saillard for their help.

## 8. References

[1] Luigi Lancieri; The concept of informational ecology or the interest of information reuse in the company. In the proceeding of 3 th. International Conference on Enterprise Information System (IEEE, AAAI) Portugal 200. Availlable olso on http://www.ensicaen.ismra.fr/lancieri
[2] Joel de Rosnay; L'homme symbiotique (the symbiotic man) Ed. du Seuil 1995
[3] Luigi Lancieri ; Memory and forgetfulness, two complementary machanism to caracterize the various actors of the Internet and their interactions; PhD Thesis, university of Caen, 2000.
[4] Web site of the French National cache network http://www.serveurs-nationaux.jussieu.fr/cache/mrtg /
[5] Nicolas Saillard Draft PhD Report/ratio; Optimization of architectures of réplications by the characterization of the traffic.
[6] Philippe Rochat, Stuart Thompson; Proxy Caching based on object location considering semantic usage; on proceedings of Web caching workshop 1999.
[7] LSAM project Web Site; http://www.isi.edu/lsam
[8] James Gwertzman; Autonomous replication; Senior thesis Harvard university 1995;
[9] Azer Bestravos; Middleware support for datamining and knowledge discovery in large scale distributed systems; In proceedings of ACM SGMOG' 96 Datamining Workshop.
[10] Bruce Edmonds, Leor Gruendlinger, Francis Heylighen; Supporting Collective Intelligence on the Web: design, implementation and test of a self-organizing collaborative knowledge system; Evolutionary Collaborative Knowledge Project for FET Open Proposal; http://www.cpm.mmu.ac.uk/~bruce/bsi/
[11] Keiichi Nakata, Angi Voss, Marcus Juhnke and Thomas Kreifelts; Collaborative Concept Extraction from Documents German National; In proceedings of the 2nd Conference on practical aspects of knowledge management (PAKM98) Oct 1998.
[12] Terveen, L.G., and Hill, W.C. Evaluating Emergent Collaboration on the Web, in Proceedings of CSCW'98 (Seattle WA, November 1998), ACM Press.
[13] Heylighen, Francis; Collective Intelligence and its Implementation on the Web: Algorithms to Develop a Collective Mental Map.Computational & Mathematical Organization Theory. Vol. 5, no. 3, pp. 253-280.
[14] Octopus corporation Web site http://www.octopus.com/
[15] NewsHub corporation Web Site http://newshub.com/
[16] net2one corporation Web Site http://net2one.com/
[17] Moreover corporation Web Site http://w.moreover.com/
[18] Snippets corporation Web Site http://www.snippets.com/
[19] OnePage corporation Web Site http://www.onepage.com/faq.html
[20] http://www.informationweek.com/story/IWK20001117S0003
[21] Yodlee corporation Web Site http://www.yodlee.com
[22] Akamai corporation Web Site ; http://www.akamai.com
[23] Network Appliance corporation Web Site ; http://www.netapp.com
[24] RFC 2616 ; HyperText Transfer Protocol ; available on W3C Web Site; http://www.w3c.org
[25] Steve Lawrence, Lee Gilles; Accessibility and distribution of information on the Web; http://www.wwwmetrics.com
[26] The communication, state of the art; Collective book directed by P. Cabin; Human sciences editions (French)
[27] David Wolpert and Kagan Tumer, An Introduction to Collective Intelligence,Tech Report NASA-ARC-IC-99-63. (A shorter version of this paper is to
Appear in: Jeffrey M. Bradshaw, editor, Handbook of Agent Technology, AAAI Press/MIT Press, 1999). Availlable at http://ic-www.arc.nasa.gov/ic/projects/collective-intelligence.html