

Intelligent Service for Aggregation of Real Estate Market Offers

Lanin V., Nesterov R.

Department of Business Informatics
National Research University Higher School of Economics
Perm, Russia
lanin@perm.ru; mistika93@mail.ru

Osotova T.

Computer science department
Perm State National Research University
Perm, Russia
hvestya@gmail.com

Abstract – This article contains the implementation description of a real estate market offers aggregator service. Advertisement analysis is made with the aid of ontologies. A set of ontologies to describe specific websites can be extended, so the aggregator can be used for many diverse resources.

Keywords – intelligent service; real estate; ontology

I. INTRODUCTION

Real estate agents constantly analyze different information flows, so intellectual analysis of real estate market offers and monitoring services are required for their efficient work. Most of this information is semistructured and in this case conventional processing is time-consuming. Real estate information resources are topical Internet resources, newspapers and special databases.

Information aggregation and structuring tasks are increasingly timely. Apart from that, it is necessary to address information duplication and inconsistency search tasks. Semistructured information and its heterogeneous resources implies application of artificial intelligence means: text mining, Semantic Web technologies and multi-agent technologies.

Our solution is to develop intelligent service to accumulate information on real estate market offers from different resources in a single database.

II. REAL ESTATE MARKET OFFERS AGGREGATORS CLASSIFICATION

The term “Aggregator” is used to describe Internet resources and services accumulating existing real estate market offers. Database completeness, data timeliness and fidelity, search and filtering capabilities and access price are the main features of aggregators [3].

Existing resources can be classified in two ways: by database areal coverage and by the way of organizing customer relations. According to the first classification resources can be divided into two groups: global ones based on a well-known web portal platform (“Yandex.Nedvizhymost” [2]) and local ones related to the regional real estate business projects. According to the second classification resources can be divided into following groups: on-line bulletin boards, electronic

versions of free advertisements newspapers, multilisting systems, information portals and meta-aggregators.

One of the first to appear was on-line bulletin board. It is usually chargeless and topically organized database. Bulletin board is arranged as a website, where anyone can place an advertisement and visitors can read it. According to experts’ opinions, on-line bulletin boards, created simultaneously with the growth of a real estate market, establish themselves a lot more firmly. New bulletin boards projects do not approach a market because they require significant capital and financial inputs. Specialists call bulletin boards a “dirty” database, i.e. disorganized and almost unregulated. Nowadays boards prevent real estate market from proper functioning, because, generally speaking, bulletin boards creators are not interested in information structuring and quality enhancing as well as in information exchange cost reduction of real estate market participants.

Electronic versions of free advertisements newspapers are also one of the core information aggregators. For instance, they include “Iz ruk v ruki” website. According to experts’ opinions, the main advantage, that allows this kind of resources to take the lead in their market segments, is that it combines newspaper concept with its electronic version. That is why, non-Internet users can also be involved, so much larger market coverage will be provided.

Among real estate brokers the most popular and in-demand kind of resources is multilisting systems. The major difference between real estate market aggregators in Russia and in western countries that in latter ones portals are owned by non-governmental organizations. Multilisting is a basis used by all market participants. For example, National Association of Realtors in USA owns the world largest real estate information aggregator “Realtor.com”. At present Russia has no global portals that would aggregate information on all real estate market offers. Commercial portals created as business projects in different Russian regions occupy this niche market.

Real estate information portals or specific (customer-oriented) websites are the most widespread aggregators of real estate information on the Internet. They are the projects that can capture its audience by having a database and providing information uniqueness, convenient delivery, wide range of analytical services, specific positioning and target audience

choosing means. Experts say that these portals appeal to users because they offer more specific information: news, analysis and wider range of search filters. Services of real estate information portals are more convenient than ones of multi-purpose websites because such portals are designed especially for keeping real estate information.

Social networks also can be called information aggregators. It is a CRM-direction that implies a step when a customer interacts with an agent. Nonetheless, today the distance between website-aggregators and social networks is shortening from the point of view on common features and applied services. In western countries this technologies are long since popular and mainly because due to Web 2.0 technologies a website visitor is becoming an information co-author and increase its reliance among other society members.

Meta-aggregator is a system accumulating real estate offers from several resources. The examples are “Skaner Nedvizhymosti” (rent-scaner.ru), “Choister” (choister.ru) and BLDR (bldr.ru). These resources offer extra services like intellectual advertisement search placed only by an owner, not by a broker.

III. DESCRIPTION OF SERVICE IMPLEMENTATION

A. Service architecture

To address automatic population of a real estate items database in the context of project on creating a real estate agency automation system an intelligent service was implemented. It extracts information on real estate items from unstructured advertisements placed on different resources. The solution is based upon an ontological approach. The general architecture of the implemented service is shown in fig. 1.

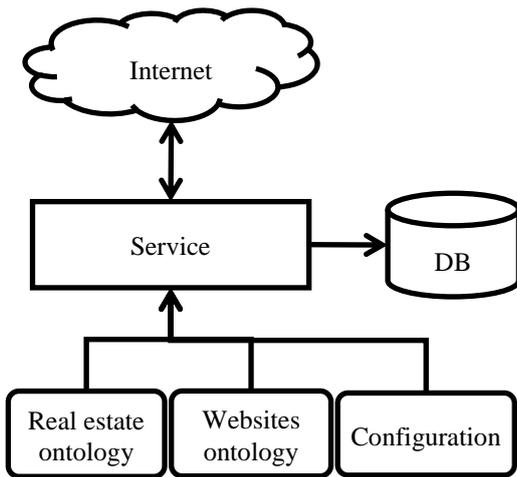


Fig. 1. Common service architecture

B. Service work layout

The general service work layout is shown in fig. 2.

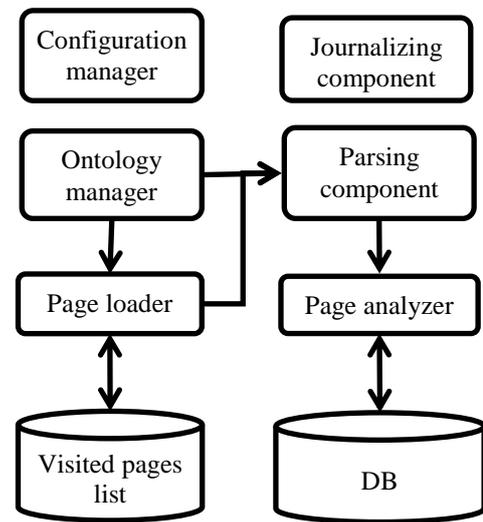


Fig. 2. Main modules of the service

Journalizing component makes a record of a service work (functioning). This record is used for service monitoring and debugging.

Configuration manager gives access to service settings and if necessary dynamically configures the service.

Ontology manager operates with ontological resources.

Page loader creates a local copy of a page and exercises its preprocessing. Information on visited pages is put into a special database. Due to this, during one loading session the loader will not visit the same pages; thus, it allows to improve service work. On basis of real estate websites ontology the loader extract information from the page. In this fashion page parser will have preprocessed text of a real estate advertisement, from which it extracts knowledge using real estate items ontology. Then this knowledge is unified (e.g. floor space can be converted to square meters).

Page analyzer makes an inference using real estate items ontology and captured knowledge as well as it checks several additional heuristics, after that it forms an object to be put into a correspond database.

C. Real estate websites ontology

Real estate websites ontology keeps specific websites settings. We are interested in keeping following parameters:

- 1) Position on a page, where the information will be found most likely and a description to have a title of this information;
- 2) Position on a page, where useful references can be found;
- 3) Description of filters to toss out “garbage” references for our service;
- 4) “Page turning” mechanism settings (more details on this are given below).

D. Real estate items ontology and regular expressions

Real estate items ontology keeps general domain concepts and their interconnections.

While parsing pages, the service attempts to “bind” specific concepts using ontology knowledge. Specific regular expressions are attached to each ontology concepts. There are two categories of regular expressions: general and website-adjusted. The latter can be used for binding only at specific websites and in general they are wrong (they allows to parse specific wordings used on a website more effectively). General regular expressions come into action in general cases. Firstly, binding of specific concepts is implemented using website-adjusted regular expressions and then in case of failure using general ones.

A regular expression consists of two components: ones to show that coincidence was found and ones to show erroneous binding. For example, “telephone” concept is binding (i.e. there is phone line), however advertisements often give a placer phone number. The second type of components used to identify a situation when it is said about different concept.

Apart from that, while extracting knowledge from text specific figures are also being bound: e.g. “Flat floor space” or “Focal person phone number”. The general structure of regular expressions is the same with the above, but additionally there are several logic parts used to convert figures to a single system. For example, if the price in advertisement were in rubles per Are, the service will convert it to rubles per square meter.

E. “Page turning” mechanism

While analyzing real estate items advertisements websites structure, it was identified that there are often lists containing advertisements references. A website has a plenty of advertisements and they are placed on different pages, that is why page crossing is implemented with navigation buttons.

We develop a “page turning” mechanism to exercise a sequential page crossing. Settings required for it are kept in real estate websites ontology and custom for each website.

It stands to mention that reference click-through, when a part of list is loaded with JavaScript, is a bit difficult to process. It was addressed by using special classes.

F. Service settings and load list

Service settings include parameters responsible for service functioning. There are following parameters:

- 1) Load list path with addresses that will be scanned by the service.
- 2) A path for saving loaded pages;
- 3) Time lapse, in that the service will resume work (service functioning can be stopped when it scanned all addresses in a load list);

4) “Websites scanning depth”, i.e. maximum path length to be scanned by the service;

5) Flag showing whether to go to third-party websites in case of the “in-depth” search.

G. “Page turning” mechanism

While analyzing real estate items advertisements websites structure, it was identified that there are often lists containing advertisements references. A website has a plenty of advertisements and they are placed on different pages, that is why page crossing is implemented with navigation buttons.

We develop a “page turning” mechanism to exercise a sequential page crossing. Settings required for it are kept in real estate websites ontology and custom for each website.

It stands to mentions that reference click-through, when a part of list is loaded with JavaScript, is slightly difficult to process. It was addressed by using special classes.

H. Programming and software tools

The service was developed using Microsoft Visual Studio 2010 and C# programming language. Ontology was developed with the aid of Protégé ontology editor. Also, HtmlAgilityPack (for html-pages parsing) and OwlDotNetApi (for reading ontologies from a file) libraries were used.

IV. BENCHMARKING

The service demonstrates rather high accuracy rates. Approximately 97 per cent of all advertisements are recognized in an adequate way. In 93 per cent of the time advertisement attributes are recognized precisely. Recognition accuracy can be improved with the aid of adjusting ontology to the specific representation of an advertisement. Besides, logging component includes analytical tools to find a reason for a fail correlation and to recommend on required ontology settings.

V. CONCLUSION

In this paper we described the architecture and peculiar implementation properties of the service aggregating the real estate market offers. At the moment the pilot system of the service is implemented. One of the core features of this service is that it can be adjusted to analysis of new resources without changing program code; configuration is only about ontology editing. Also, in the context of this project and on the basis of the information kept in the system, we intend to develop an expert system on selection and estimating of real estate items.

REFERENCES

- [1] Segaran T., Evans C., Taylor J. Programming the Semantic Web, O'Reilly Media, 2009.
- [2] Что такое Яндекс.Недвижимость <http://help.yandex.ru/realty/>
- [3] Недвижимость online: агрегаторы <http://media-office.ru/?go=2082914&pass=f79e9c77f077cf1d060a615834c3c2d1>