



JOB VACANCIES GATHERING ENGINE DEVELOPMENT



E. ELOKHOV, R. NETSEROV, I. OKULOVA NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, PERM

INTRODUCTION

Nowadays applicants and employers communicate over various systems such as professional networks or headhunter sites. From examining the findings, limitations of existing job search applications restrict opportunities for analysis. In order to assess how trustworthy an employer is, it seems that there is no automated method of analyzing employer characteristics such as personnel turnover and salary level, to name but a few.

One of the reasons why this issue remains complicated is information distribution. Vacancies are posted in numerous websites so that it complicates data processing. Even though automated aggregation systems simplify this procedure, the algorithms are executed inaccurately.

EXISTING APPROACHES

In Russia the most applicable automated job search service is **Yandex.Rabota** [1]. It can be characterized as follows:

1. Numerous URLs of vacancies from various head hunter websites.
2. Specific Yandex Vacancy Language (YVL) exploitation to establish opportunities for data processing.
3. Requirements for head hunter website the provide proper vacancies aggregation by means of YVL.

Indeed.com [2] is a search engine for jobs that drives job searchers directly to jobs on corporate career websites, employee recruiting job boards, online newspapers, blogs, and association websites. It has following features:

1. Thousands of different recruiting websites used by employers indexing.
2. Salary search opportunities including national salary trends building.
3. Engine does not cover Russian recruiting websites.

PROPOSED APPROACH

By examining existing approaches, some complexity of detail analysis was found out. There is no possibility to use automated aggregators to propose new analysis due to the peculiarity of data presentation, while current data visualization and studies conducted, for example, by Yandex or Indeed groups do give insufficiently amount of information. This research is intended for minimization of lack of analyzed information.

There exists the intelligent service, which was developed to aggregate and analyze real estate market offers. Thorough implementation description is in. Potentially, this service can be configured and adjusted for other domains because ontologies are used. The general implementation architecture is presented in Fig. 1. Ontologies employed were developed with Protégé 4.3.

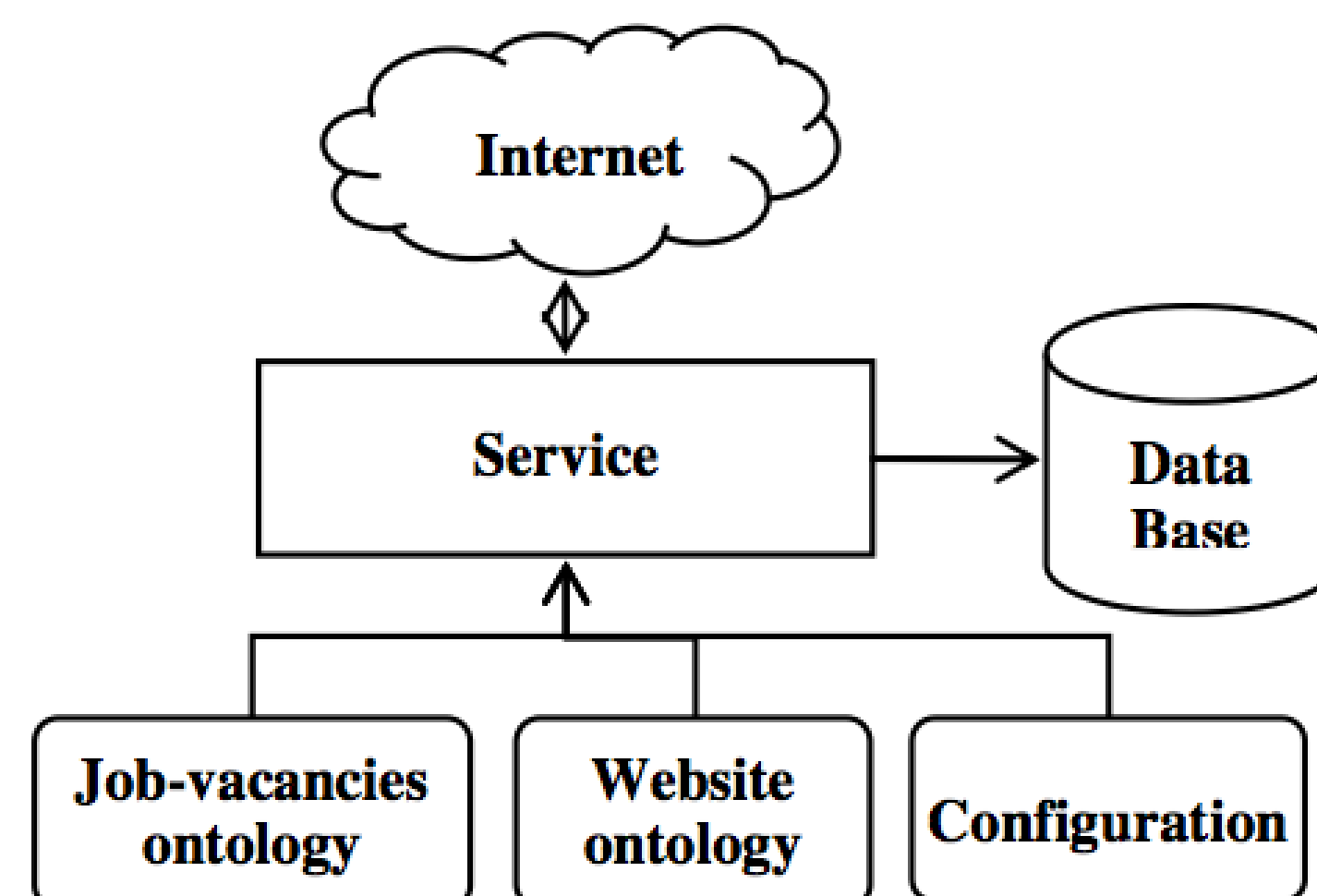


Fig.1. General architecture

An ontology of job-vacancies (domain ontology) keeps domain concepts and specific regular expressions attached to them. Website adjusted regular expression are used for binding concepts only at specific websites, while general regular expressions activate in general cases.

Sources ontology will be derived based on html page code analysis. Data properties are going to be put into websites ontology.

The process of domain ontology creation can be semiautomated or unautomated. There is a number of methodologies for semi-automated ontology building [4]. In our case, there is no need to develop and employ complicated algorithms.

Microformats can be applied to extract domain concepts and interconnections to evaluate them further. There is a number of microformats for our domain, such as [5], [6].

We do not develop here a new methodology, but we try to define data gathering and ontology enrichment life cycle as a process composed of four main steps. The process is depicted in Fig. 2.

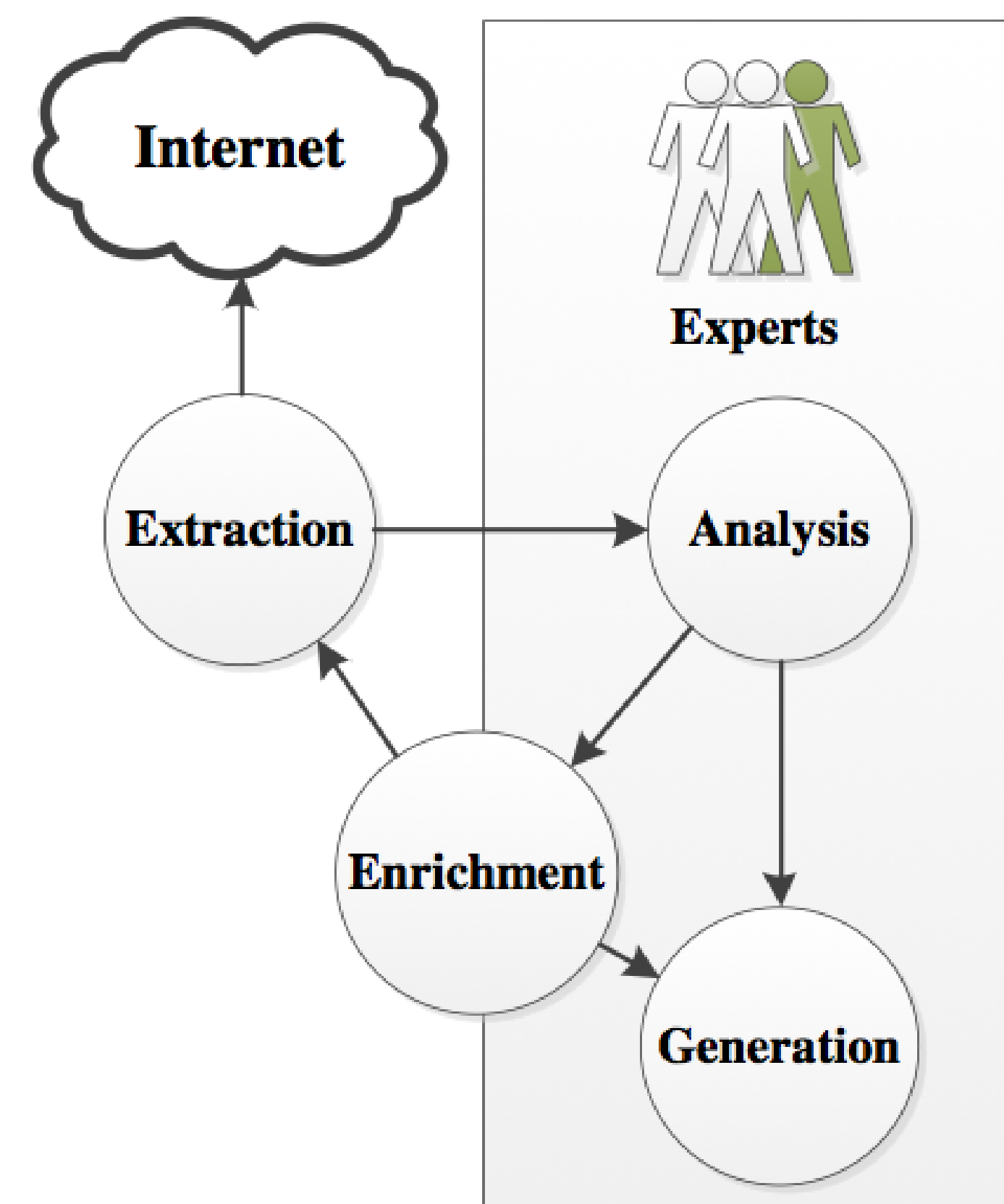


Fig.2. Ontology enrichment steps

Extraction: information acquisition needed for ontology generating (classes, attributes and relations).

Analysis implies accuracy evaluation of extract concepts at the previous stage performed by a group of experts.

Generation: basic concepts and relations creation or enriching the existing ontology.

Enrichment: an ontology is not a static description of a domain, but with the evolution of job vacancies variety, the ontology may also require some changes. The number of individuals can be added or modified. So creating new individuals is a possible way to enrich an existing ontology.

CONCLUSION

The main conclusion to be drawn is that the approach to address the information gathering problem to analyze and evaluate employer characteristics is proposed.

Proposed method will be implemented as a special service for the information source generation and a data analysis tool for the development of competitive advantage. More powerful analysis tool will enable to provide more qualitative and more useful service.

REFERENCES

1. Yandex.Rabota - vacancies [Online]. Available: <http://rabota.yandex.ru>.
2. Indeed.com - one search. all jobs [Online]. Available: <http://www.indeed.com>
3. V. Lanin, R. Nesterov and T. Osotova, "Intelligent Service for Aggregation of Real Estate Market Offers," in Proceedings of the 7th SYRCoSE, Kazan, Russia, 2013, pp. 136-138.
4. I. Bedini and B. Nguyen, "Automatic Ontology Generation: State of the Art," in University of Versailles Technical report, 2007, pp. 3-4.
5. A listing that describes a job opening in a certain organization [Online]. Available: <http://schema.org/JobPosting>.
6. Job List Brainstorming [Online]. Available: <http://microformats.org/wiki/job-listing-brainstorming>

CONTACTS

Here are the authors' e-mails to contact:

- **Eugenii Elokho**
eugene.yelokhov@gmail.com
- **Roman Nesterov**
ranesterov@edu.hse.ru
- **Irina Okulova**
ir.okulova@gmail.com

Acknowledgements: The reported study was supported by RFBR, research project No. 14-07-31273 mol_a.