



Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Structuring visual exploratory analysis of skill demand

A.-S. Dadzie^{a,*}, E.M. Sibarani^{b,c}, I. Novalija^d, S. Scerri^{b,e}^a *KMi, The Open University, Milton Keynes, UK*^b *University of Bonn, Bonn, Germany*^c *Institut Teknologi Del, Toba Samosir, Indonesia*^d *Institut Jožef Stefan, Ljubljana, Slovenia*^e *Fraunhofer IAIS, Bonn, Germany*

ARTICLE INFO

Article history:

Received 18 February 2017

Received in revised form 17 November 2017

Accepted 11 December 2017

Available online xxx

MSC:

00-01

99-00

Keywords:

Domain modeling

Knowledge discovery

Visual exploration

Ontology-guided visual analytics

Trend identification

Demand analysis

ABSTRACT

The analysis of increasingly large and diverse data for meaningful interpretation and question answering is handicapped by human cognitive limitations. Consequently, semi-automatic abstraction of complex data within structured information spaces becomes increasingly important, if its knowledge content is to support intuitive, exploratory discovery. Exploration of skill demand is an area where regularly updated, multi-dimensional data may be exploited to assess capability within the workforce to manage the demands of the modern, technology- and data-driven economy. The knowledge derived may be employed by skilled practitioners in defining career pathways, to identify where, when and how to update their skillsets in line with advancing technology and changing work demands. This same knowledge may also be used to identify the combination of skills essential in recruiting for new roles. To address the challenges inherent in exploring the complex, heterogeneous, dynamic data that feeds into such applications, we investigate the use of an ontology to guide structuring of the information space, to allow individuals and institutions to interactively explore and interpret the dynamic skill demand landscape for their specific needs. As a test case we consider the relatively new and highly dynamic field of *Data Science*, where insightful, exploratory data analysis and knowledge discovery are critical. We employ context-driven and task-centred scenarios to explore our research questions and guide iterative design, development and formative evaluation of our ontology-driven, visual exploratory discovery and analysis approach, to measure where it adds value to users' analytical activity. Our findings reinforce the potential in our approach, and point us to future paths to build on.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Exploratory discovery and detailed analysis of complex, unbounded scenarios are fundamental components of working and living in today's technology- and data-rich world. Obtaining a good understanding of such scenarios typically involves analysis of complex and/or large amounts of data. Such analysis often requires linking to other related datasets that provide context to fill in blanks in the core dataset(s) and aid understanding of the complete scenario. However, human cognitive limits make it easy to get lost in data as size, complexity and noise grow, hampering data exploration and exhaustive, in-depth analysis.

Effectively designed visualisation, by taking advantage of advanced human perception to recognise patterns and trends within

data, reduces cognitive load while aiding the user to obtain informative overviews and recognise and explore regions of interest (ROIs) in detail [1,2]. Ontologies provide a tool for structuring information spaces; in interactive visualisation this structure may be used to aid navigation through, querying and interpretation of underlying data content. Further, employing a uniform, underlying data structure supports representation from multiple, coordinated perspectives [3–5]. Ontology-guided visual analysis allows the strengths of each approach to complement the other, to enable intuitive yet structured, in-depth yet exhaustive analysis, increasing ability to obtain insight into complex data.

1.1. Motivation

The overall aim of this study is to determine the demand for jobs, and at a higher level of granularity, skills specific to a domain or industry sector and additional non-domain skills required to fill a job role. We aim to build intuitive representations of skill demand, that allow target end users to obtain a good understanding

* Corresponding author.

E-mail addresses: aba-sah.dadzie@open.ac.uk (A.-S. Dadzie),

sibarani@iai.uni-bonn.de (E.M. Sibarani), inna.novalija@ijs.si (I. Novalija),

scerri@iai.uni-bonn.de (S. Scerri).

<https://doi.org/10.1016/j.websem.2017.12.004>

1570-8268/© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of the demand landscape and how this bigger picture changes with location and time.

To meet our goals we scan job adverts using keywords corresponding to skills required to fulfil typical and specialised job roles in a domain. To allow us to evaluate our approach and also support in-depth analysis, we restrict our initial exploration to the *Data Science* domain and Europe. For completeness for this domain we capture a broad range of technical skills covering *data literacy*, *knowledge acquisition*, basic to advanced *exploration* and sophisticated *analysis*, and the *soft skills* needed to *communicate* the results of analysis within the technical domain and to a wider audience, using, e.g., *storytelling* and *data journalism*.

To ensure our analysis and results are reusable and extensible beyond this focus, we formalise our requirements for data capture, exploration and analysis using a set of ontologies describing skill demand and additional, relevant metadata, including target end user types (ranging from domain and technical experts to the interested but non-expert public); temporal attributes (dates and date ranges); and geographical location (at precision levels of country, city/town and latitude/longitude). We present the resulting knowledge framework as an upper level ontology that provides the backbone for structuring our analysis. This allows us to reuse the ontology for the more general case of skill demand analysis across any named domain and location. By capturing formally, also, the requirements of our target end users we obtain a framework that supports, further, the design and development of intuitive, interactive tools for task-driven, exploratory knowledge discovery and guided analysis.

To guide our study we explore the following data-driven research questions (RQs), with an aim to propose practicable solutions for the challenges that arise in the exploration of skill demand:

- RQ1:** what is the demand landscape for data scientists across the European Union (EU)?
- RQ2:** to what extent is demand matched by competence, as measured by skills in the existing workforce? Does this change when restricted to a specified location, time period, domain or industry sector?
- RQ3:** what support exists for a given individual – currently in the workforce or a trainee – seeking to identify and fill their “ideal” or desired job role?

Exploring these questions requires detailed analysis of job demand, using data from multiple, heterogeneous sources, with variation in attributes specified, terminology used and level of detail in job role descriptions. We employ use case scenarios describing our initial focus and target end user types to analyse the fuller set of challenges these RQs raise. Our findings feed into design and development of support for these, and ultimately, other end users exploring skill demand. For instance, successfully answering **RQ1** should allow us to provide intuitive support for, say, the general public, and therefore, a policy-maker at state level to retrieve information that answers the question:

How much movement would be seen within the EU for highly skilled statisticians seeking employment for the two-year period from mid 2016? Would access to more advanced, tertiary-level courses in object-oriented programming influence this migration?

Answering RQs 2 and 3 requires assessment also of existing expertise and the range of support in place for training and upskilling. This would allow a job seeker to answer questions such as:

What other skills would I need to find a job that allows me to combine my interest in journalism with the expertise I developed in my masters year internship building an iPad app to help pensioners keep track of their finances?

We will discuss the user- and task-centred process we follow toward answering such questions, and employ the working tools that support the exploratory discovery and sense-making tasks required to do so to review also the skill demand datastore we are building. This is a necessary precursor to designing intuitive support that will allow our varied target end users to independently explore and carry out further exploration and detailed analysis as needed.

Section 2 details one of a set of use cases developed to elicit end user requirements and illustrate the process we follow to generate an interactive landscape of skill demand. Section 3 reviews related work in job and skill demand analysis, and the value in visual, ontology-guided analysis for scenarios such as those we explore. We introduce our study methodology in Section 4, and discuss in Section 5 findings from our exploration of the initial dataset collected to describe skill demand. We feed these findings into a review of our initial requirements (Section 5.1), and formalise the outcomes in the *Skills and Recruitment Ontology – SARO* (Section 6). Section 7 describes the information extraction process that feeds into building the data store structured according to the requirements specified in SARO. We describe formative evaluation of our framework in Section 8 and illustrate, using the paths that participants followed to complete set tasks, tool design and functionality toward intuitive analysis of skill demand. We revisit our research questions and the use case scenarios in Section 9, to assess the extent to which our process enables target users to answer their questions. We also discuss, in this and within the context of each section, where future work will bring us closer to our overall goal. We conclude the paper in Section 10.

2. Use case scenarios

Recent technological trends show record volumes of data being generated and captured [6–8]. While this has the potential to contribute added value across the European economy and worldwide, efficiency of extraction and effectiveness in the use of its content impacts actual value recouped from this knowledge source. This modern challenge, due to and that can only be resolved through effective use of technology, was a key factor in the selection of our use case and initial application – to measure the gap between market demand for skilled analysts of *big data* and practitioners with the expertise necessary to meet this demand.

Our study methodology involves the exploration of skill demand and supply scenarios that consider the information-seeking and analysis requirements of the five target end user types identified in the early stages of the study [9], the: (1) policy- or decision-maker, (2) educator or trainer, (3) recruitment agent, (4) practitioner and the trainee or (skilled) job seeker and (5) the interested public. Based on the initial exploration of the requirements of our targets we created a task-based questionnaire that we use to guide our user-centred design (UCD) process, following an iterative cycle of design, development and evaluation. This paper extends the findings from this study [9] and reviews the initial requirements, to feed into the current exploratory and analytical stage. We focus here on the *practitioner* and the *trainee/job seeker* aspiring to this role – technical end users with data science-oriented backgrounds, and also non-technical users with good knowledge of the requirements for working with big data. Where relevant and for completeness, we highlight where requirements and our findings overlap with those of the other user types, as we explore in depth the data, and in an iterative UCD cycle, design for user-focused exploration and analysis tasks in line with users’ interim and ultimate information seeking goals.

Download English Version:

<https://daneshyari.com/en/article/6950441>

Download Persian Version:

<https://daneshyari.com/article/6950441>

[Daneshyari.com](https://daneshyari.com)