



# Competence Mining to Improve Training Programs

Ildikó Szabó<sup>(✉)</sup> , Katalin Ternai , and Szabina Fodor 

Corvinus University of Budapest, Fővám tér 13-15, Budapest 1093, Hungary  
{ildiko.szabo2,katalin.ternai,  
szabina.fodor}@uni-corvinus.hu

**Abstract.** Analysing of competence and skill shortage or surpluses is essential for educational institutes to prepare their students for satisfying labour market needs in time and comprehensively. Currently, changes in labour market needs are influenced by not just economical but also technological factors. ICT and digitalization play key roles in transformations of business processes including employees' competences in executing these processes smoothly and effectively. Our research goal is to develop a competence mining method to identify and extract competences needed to fill job vacancies. Based on this new information the educational programs can be refined. This paper presents how to use business process models to extract competences from job vacancies and how this method evolved in time and what its contribution is to the training development based on learning outcome. Competence concept has a crucial role in this method, but it is defined on a broad scale that causes terminological diversity.

**Keywords:** Training program · Competence · Text mining · Business process model · Ontology transformation

## 1 Introduction

Educational institutes must meet different expectations. Globalization puts pressure on institutes to train students who can adapt to different culture and environments during their whole lifetime. Machines are unable to replicate judgement, empathy, persuasion, the ability to collaborate and communicate, and be flexible, adaptable, and resilient, though more AI change is coming [1]. These skills should be possessed by students. In 2018, JISC<sup>1</sup> in the UK surveyed 22,000 university students and while more than 80% of them felt that digital skills will be important for their chosen career, only half believed that their courses were preparing them well for the digital workplace [2]. The Universities need to start to develop an education 4.0 program in line with the 4th industrial revolution. The education 4.0 program should aim at providing graduates with capabilities and competencies required by the digital-driven industry [3]. The newest industrial revolution is boosted by technologies that creates new requirements. This transformation induced by Industry 4.0 requires continuous education processes

<sup>1</sup> Formerly the Joint Information Systems Committee (<https://www.jisc.ac.uk>).

performed by humans, educational programs fitting to information technologies, providing multidimensional mind development and deep understanding of subjects. These programs have to train such students who can learn and like analysing and discussing problems [4]. Higher educational institutions need to address needs of the labour market and this should be linked to research issues.

For the point of view to this study, monitoring competence needs of labour market is essential for educational institutes. Job vacancies published on a job portal aggregating posting from many sites (like indeed.com) are appropriate sources for ensuring this activity. These job ads are numerous and time varying, hence reflect dynamically changing needs well. The goal of our research is to develop a competence mining method to identify and extract competences needed to fill job vacancies. Based on this new information the educational programs can be refined. This paper presents how this method evolved in time and what its contribution is to the training development based on learning outcome. Competence concept has a crucial role in this method, but it is defined on a broad scale that causes terminological diversity. Section 2 presents some definitions and shows the role of this concept in training development. It clarifies this topic from the point of view of our method. Section 3 sheds light on some current research to emphasize our novelty in this field. Section 4 presents how competences can be mined from 192 warehouse manager jobs published the one of top job portals (indeed.com) in UK with using general text mining method and our process-based text mining method. Finally, it presents a competence dashboard for educators to improve their training programs. Conclusions and future works are drawn in Sect. 5.

## 2 Outcome-Based Training Development

Competence expresses “What is a human capable of?” [5]. Hecklau et al. [6] classified future competencies into four categories based on macroenvironmental analysis. Technical competencies are knowledge, skills and abilities related to work. Personal competencies are motivations and attitudes of people. Social competencies are abilities to cooperate, communicate with other people. Methodological competences are to support decision making and problem solving. Wikle & Fagin [7] distinguished hard/technical and soft skills as competences. Hard skills are learned or professional competencies, soft skills are generic ones. Lippman et al. [8 pp. 4] defined soft skills as “a broad set of skills, competencies, behaviours, attitudes, and personal qualities that enable people to effectively navigate their environment, work well with other and perform well”. In the previous decades, there was a paradigm shift in education that put students in focus instead of teachers. European Qualification Framework represents this endeavour well. It advises to focus on what “students are expected to achieve and how they should demonstrate what they have learned” [9]. The emphasis was shifted from the learning output to learning outcome possessed by students. Conceptualizing learning outcomes helps in designing curriculum including teaching and learning activities, assessment methods and in creating transparent training programs. In Gagné’s theory, the planning of education programs starts with the identification of learning outcomes followed by the construction of the task analyses – or in other words, the learning hierarchy – that are responsible for execution of measurable

activities [10]. Technical/methodological competency and hard skill are very similar concepts. Social or personal competencies can be considered as soft skills. The original goal of Benjamin S. Bloom was to elaborate a better way to compare results of various training programs and test methodologies in 1956. His method provided guidelines for elaborating various training programs subordinated to learning outcomes. Terminology, structural changes, and changes in emphasis were made in its revised version. The Revised Taxonomy contains six categories: Remembering, Understanding, Applying, Analysing, Evaluating, Creating. Specific verbs are selected to describe these categories [11]. In summary, it was presented that outcome-based training development requires well-defined and measurable learning outcomes which are expressed by competences. Revised Bloom taxonomy presents six cognitive levels of complex thinking that are measurable by tasks achieved by students. Verbs describing these levels can be applied to extract competences or skills from job ads (see Sect. 4). Our research goal is to develop a competence mining method to identify, extract competences needed to fill job vacancies in order that training programs can be adjusted to this competence set. Business process model serves as a basis of this text mining process because it contains tasks which activity part can be connected to the revised Bloom taxonomy. It also provides additional information to understand the complexity of these tasks and other related factors. Hard skills or technical/methodological competencies are connected to tasks primarily that is why our purpose is to mine these competences instead of soft skills or social/personal competencies.

### 3 Related Work

OECD<sup>2</sup> Skills for Jobs database<sup>3</sup> is designed to measure skill shortage and surplus. Indicators to reveal these discrepancies were created based on five sub-indices: wage growth, employment growth, hours worked growth, unemployment rate, under-qualification growth. These macro-economic indicators are to estimate changes in occupation groups and related skill sets [12]. CEDEFOP<sup>4</sup> Skill Forecast uses quantitative methods to forecast future trends in sector, occupations, and qualifications. Skill-OVATE is an online vacancy analysis tool for Europe. It provides insight into skills and jobs requested by employers. Data are fetched from job portals, employer's portals etc. but within a given time and not dynamically [13]. Bakhshi et al. [14] used occupations ranked by experts to create machine learning method for analysing future competence needs. Skill market gap analysis of SMART system is to identify gaps between competences can be acquired by a tourism-specific training program and labour market needs in tourism industry. This system also processes job ads but uses domain ontology and not process ontology to identify competences in them [15]. Skill gap analysis is a hot topic currently, hence the above-mentioned researches are just few examples among different projects and initiatives. However, they are different in

<sup>2</sup> Organisation for Economic Co-operation and Development.

<sup>3</sup> <https://www.oecdskillsforjobsdatabase.org>.

<sup>4</sup> Centre Européen pour le Développement de la Formation Professionnelle.

methods: quantitative and/or text mining, machine learning methods are applied. Our approach distinguishes from that based on the fact that it provides a tool to monitor labour market needs and it uses process ontologies as underlying knowledge.

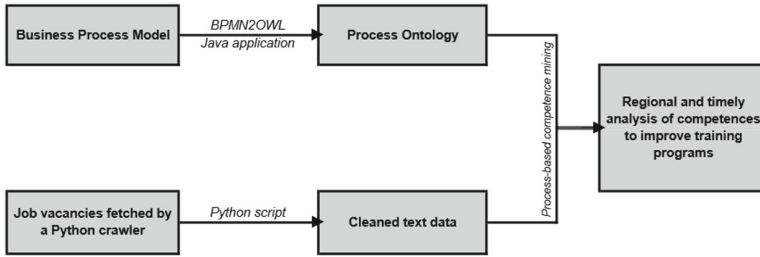
#### 4 Contribution: Process-Based Text Mining

Our main purpose is to analyse competences are required to fill a position. Several positions can be digitally transformed due to technological innovations (like self-driving forklift or mobile app). A reference business process model considering these innovations holds background information to process job vacancies in meaningful manner. The process-based text mining process is illustrated by Fig. 1. At the beginning, the structure of a process model is designed (task as process step, role as job role and required skill to execute this task by this role).

The business process model is implemented by using BOC ADONIS modelling platform<sup>5</sup>. ADONIS is a graph-structured BPM language. The ADONIS modeling platform is a business meta-modeling tool with components such as modeling, analysis, simulation, evaluation, process costing, documentation, staff management, and import-export. Its main feature is its method independence. Our approach is principally transferable to other semi-formal modelling languages. The models can be exported in the structure of ADONIS XML format. There are several parameters that can be set or defined when modelling a business process. The shell of a business process can be easily formed with activities, decision points, parallelism or merging objects, logical gateways and events. A prototypical java tool was developed to transform the business process into a process ontology in OWL format. For the mapping the conceptual models to ontology models meta-modeling approach have been used. The “conceptual model - ontology model” converter maps the Adonis model elements to the appropriate ontology elements in meta-level. The general rule used in our approach is to express each ADONIS model element as a class in the ontology and its corresponding attributes as attributes of the class.

Meanwhile a Python crawler fetches information about job vacancies from the selected job portal. Data cleaning and data prep process are executed before the text mining. The process-based text mining identifies patterns of part-of-speech tags (e.g. verb followed by noun) to get a list of expressions describing tasks (see Bloom taxonomy). The algorithm calculates the semantic distance of these expressions (e.g. create quality) from the business process elements of the process ontology (e.g. check quality). Similarity coefficients are used to do this calculation. The list of expressions is filtered by the value of the selected coefficient and by the descriptions of process elements. Remained expressions as descriptors identify process elements (like tasks) in job ads. Job ads contain information about when and where these process elements were required and by which position. An extended table is used to analyse task-related competences regionally, in time and based on positions. The theoretical background of

<sup>5</sup> BOC Group: Business Process Management with Adonis, <http://www.boc-group.com/products/adonis/en/>.



**Fig. 1.** Process-based text mining

this process has been presented in [16]. This paper presents how business process models can provide additional information in processing job vacancies versus basic text analytics method. The whole process is illustrated by the purchasing process including activities performed by warehouse managers.

#### 4.1 Business Process Model and Its Transformation

A process from logistics was selected to present the applicability of the method. The warehouse manager has several responsibilities in this process, so this profession is well applicable to illustrate our method.

The warehouse manager's activities in the process are the following ones:

- Check Inventory: Stock-taking is the physical verification of the quantities and condition of items held in the inventory or warehouse.
- Check Quality: The quality assessment will be based on predetermined requirements and standards previously set by the company.
- Unload Goods: Safely prepare, lift, position and restrain goods on a vehicle platform and then unload goods at the destination.
- Manage Goods In: Provide the correct goods, at the correct amount, place, and time.
- Record Data: Record the purchasing information into the warehouse management information system.
- Treat Scrap: Sort the scrap into recyclables and unusable waste.
- Record Report: Record the waste information into the quality management system.

For the transformation process, a prototypical software tool was developed which transforms the BPMN model into OWL format<sup>6</sup>. The resulting file contains a partial ontology including classes and individuals of the input file [16].

<sup>6</sup> The BPMN model and the transformation program are available on the GitHub (<https://github.com/szabinaf/BPM2OWL/>).

## 4.2 Job Vacancies and Data Preparation

192 warehouse manager jobs were fetched from the UK labour market<sup>7</sup> by a Python crawler into CSV file. It was discovered that the important part of job descriptions starts with specific expressions such as responsibilities, accountability, looking for, candidate and so on. The 90% of job vacancies met these expectations, but the remained ones did not contain unnecessary texts (like company description). Hence the first part of these descriptions was cut off to get the relevant part of them.

Spacy, pandas, re and text distance Python library were used. List was created from the sentences of cleaned job descriptions. Sentences with at least 20 characters were processed, lemmatized, tokenized and their part-of-speech tags were determined.

## 4.3 First Approach: Text-Mining Based on Job Vacancies

Our first approach was to extract competence information based on only data downloaded from the career website. The cleaned and tokenized job description field was used for basic text analysis.

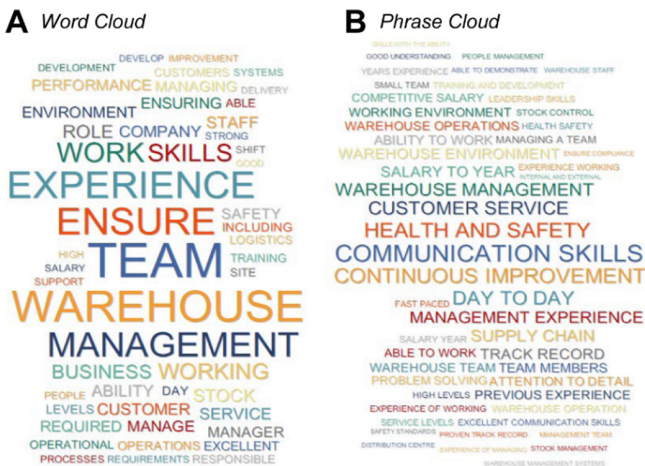
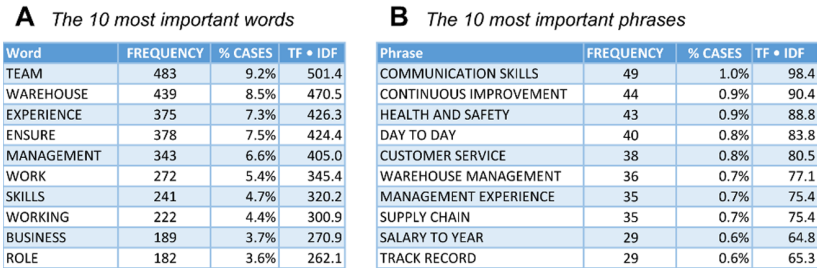


Fig. 2. Word and phrase cloud based on frequency

First, word, phrase frequency analysis was executed and word and phrase cloud was generated (see Fig. 2). From the point of gathering competence the most used words as “TEAM”, “WAREHOUSE” and “ENSURE” cannot add useful information.

<sup>7</sup> [uk.indeed.com](http://uk.indeed.com).

Looking into the most common terms like “COMMUNICATION SKILLS”, “CONTINUOUS IMPROVEMENT”, “HEALTH AND SAFETY” seem more relevant information. Analysis of the top 25% (1247 out of 4986) of used phrases provides valuable insight of needed competences such as “PROBLEM SOLVING”, “EXCELLENT COMMUNICATION SKILLS”, and “MANAGING A TEAM”.



**Fig. 3.** The 10 most important words and phrases based on TF-IDF

Figure 3 shows the most important words and phrases of job descriptions according to TF-IDF<sup>8</sup>. It provides good opportunity to find additional necessary competences like “TEAM MEMBERS”, “ATTENTION TO DETAIL”, “PROBLEM SOLVING”.

Then co-occurrence analysis was performed and different similarity coefficients (like Jaccard, Sorensen, Association strength, Adjusted Phi Coefficient) [17] were used to measure the distance between “ability/able”<sup>9</sup> keywords and words of job description field data (see Fig. 4). In groups made with different coefficients, many common words appeared like “DEMONSTRATE”, “WORK”, which confirms the importance of these words among the competencies of the warehouse manager position. It is also worth noting that among the words close to the keywords “ability” and “able” it can be found a number of verbs also used in Bloom taxonomy, such as “DEMONSTRATE” (in group Apply), “MANAGE” (in group Analyze), which confirms the effectiveness of our research. It worth mentioning that the similarity scores of text analytics are very low which indicates partial role in information.

<sup>8</sup> TF-IDF (term frequency-inverse document frequency): a statistical measure that evaluates how relevant a word is to a document in a collection of documents.

<sup>9</sup> The “ability” and “able” as keywords were used to describe the meaning of competence.

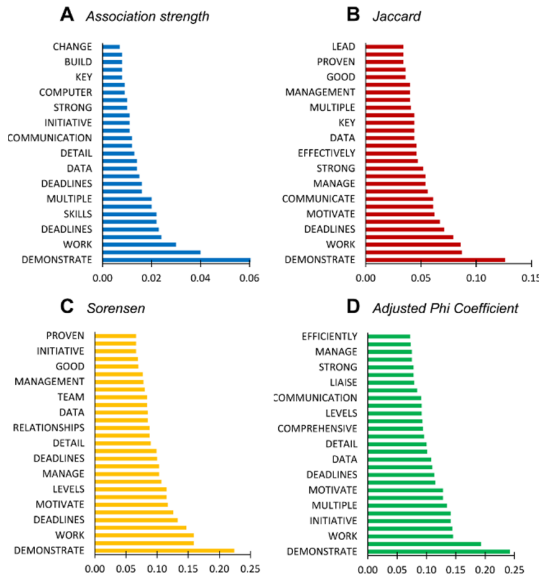


Fig. 4. 15 proximate words to “ability” and “able” based on different similarity coefficients

#### 4.4 Second Approach: Process-Based Competence Mining

Please note that the even results of basic text analytics are too general, but it underlies the usability of Bloom taxonomy. Hence expressions in the form of verb and noun were extracted from these sentences. 1652 expressions were discovered in 4989 sentences. After removing duplicates 1324 expressions left as corpus.

The process steps were extracted from the process ontology. Jaccard normalized similarity coefficient was applied to determine which verb-noun expressions from the corpus are related to our role (warehouse manager) semantically. Expressions with greater than 0.5 Jaccard points were selected. Descriptions of process steps (see in Sect. 4.1) provided background information about the basic tasks that a warehouse manager has to do: check inventory (56% of 25), manage goods in (60% of 45), unload goods (50% of 6), check quality (80% of 5), record data (0% of 16), treat scrap (0% of 19), record report (39% of 23). Please note that if business process manager does not follow the professional terminology (e.g. treat scrap) or uses too general expressions (e.g. record data), the process model does not help in extracting competences from the descriptions. Otherwise it provides several descriptors to identify tasks in texts. All 57 descriptors were searched in each job descriptions. It was a good experience that not all job ads contained one or some of them. The reason behind this is that positions usually contain more than one role (e.g. executive, responsible roles etc.).

This study focused on only warehouse manager as a specific role. If the tasks executed by this role were not emphasized in the job ads but tasks related to other role (e.g. team leader activities) were, it diminished the hit rate aggressively. Our plan is to extend our business processes with more tasks related to other roles based on this analysis and filter our job ads by roles – not just positions - more precisely. Nevertheless, this method



processed enough job ads to present how to use extracted tasks as technical/methodological competences or hard skills to improve curricula of training programs.

#### 4.5 Potential Improvements in Training Programs

All job ads have information about where and when job vacancies were created in which position and at which company. The Python crawler stores this information beside job descriptions in a table. This table is extended with new columns named the above mentioned 57 descriptors and 7 process steps. Each row of these new columns has 1 or 0 value, depending on the fact that the job description in the row contain the given descriptor or not. Columns named process steps summarize these values. This extended table is appropriate to analyse competence needs from different viewpoints (regions, time, position, company, salary etc.). A new column was added to this table to examine digitalization needs in warehouse manager positions. This column represented that system knowledge was needed to fill the position or not. Advertising companies used mainly warehouse management systems in that time (mentioned in 18% of the job ads), but other 15 systems (such as stock control, operational management, ERP system etc.) appeared in the ads as well.

Figure 5 shows the ‘*manage goods in*’ competence was required in the middle and south of United Kingdom, meanwhile system knowledge was also needed to fill warehouse manager positions there. It highlights the importance of this task and knowledge that can be taken into consideration by educational institute in this region. Additional information can be retrieved by other analyses. It can be presented timely distribution of competence needs to examine seasonal requirements. Training programs qualify students to fill given positions hence educators can be interested in seeing timely distribution of competences needs related to a specific position. Data can be drilled down in multidimensional analysis to reveal this kind of correlations. A system built on this method is capable of continually monitoring the labour market needs and providing information about competence trends and distributions. Future competence requirements can be predicated based on the trends. Programme leaders can get regional and temporal feedback to evolve their training programs with adjusting them to the actual or future needs of relevant jobs.

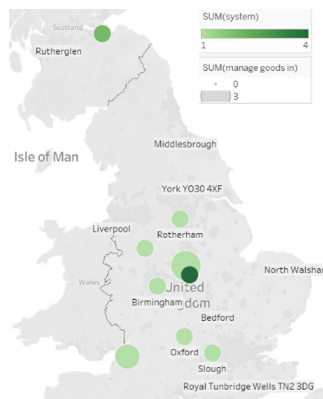


Fig. 5. The ‘manage goods in’ competence in a regional analysis

## 5 Conclusion

This paper presents how to use business process models to extract competences from job vacancies and how to use the results to improve training programs of educational institute. It illustrates through an example what competences are provided to stakeholders with using traditional text mining or process-based text mining. Comparing these methods, it revealed that additional information are required to filter the corpus or glossary, to highlight system usage information, and to manage multiple roles in positions. The process model ensures these information, because systems, executors and tasks are connected in it. In this way, process-based text mining discovers more specific competences versus traditional text mining. A tool built on this method is capable of detecting what kind of systems and transactions should be operated by the warehouse manager. Information about mass competences needs distributing regionally and timely are used to reform training programs or optimize human resource training. Stakeholders can reorganize the educational portfolios, and also manage capacities. The key limitation of this study is that it largely depends on the phrases and terms used in the process model. In the future, alternative terms will be incorporated into our model and the scope of the business process model will be extended, and the pattern sets will be expanded.

## References

1. JISC: Preparing for Education 4.0. <https://www.timeshighereducation.com/hub/jisc/p/preparing-education-40>. Accessed 2019
2. JISC: Digital skills crucial to the success of fourth industrial revolution. <https://www.jisc.ac.uk/news/digital-skills-crucial-to-the-success-of-fourth-industrial-revolution>. Accessed 2019
3. Rasika, L., Lim, F.C., Haslinda, A.: Strengths and weaknesses of education 4.0 in the higher education institution. *Int. J. Innovative Technol. Exploring Eng. (IJITEE)* **9**(2S3) (2019). ISSN: 2278-3075
4. Akgül, H.: Examining the impact of industry 4.0 on education. *J. Awareness (JoA)* **5**, 159–168 (2020)
5. Rowe, C.: Clarifying the use of competence and competency models in recruitment, assessment and staff development. *Ind. Commercial Training* **27**, 12–17 (1995)
6. Hecklau, F., Galeitzke, M., Flachs, S., Kohl, H.: Holistic approach for human resource management in industry 4.0. *Procedia CIRP* **54**, 1–6 (2016)
7. Wikle, T.A., Fagin, T.D.: Hard and soft skills in preparing GIS professionals: comparing perceptions of employers and educators. *Trans. GIS* **19**, 641–652 (2015)
8. Lippman, L.H., Ryberg, R., Carney, R., Moore, K.: Workforce connections: key “soft skills” that foster youth workforce success: toward a consensus across fields | VOCED plus, the international tertiary education and research database. Child Trends publication (2015)
9. Havnes, A., Prøitz, T.S.: Why use learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educ. Asse. Eval. Acc.* **28**(3), 205–223 (2016). <https://doi.org/10.1007/s11092-016-9243-z>
10. Prøitz, T.S.: Learning outcomes: What are they? Who defines them? When and where are they defined? *Educ. Asse. Eval. Acc.* **22**, 119–137 (2010). <https://doi.org/10.1007/s11092-010-9097-8>

11. Forehand, M.: Bloom's taxonomy. *Emerg. Perspect. Learn. Teach. Technol.* **41**, 47–56 (2010)
12. OECD: *Getting Skills Right: Skills for Jobs Indicators*. OECD (2017)
13. Cedefop: Skill forecast. <http://www.cedefop.europa.eu/en/publications-and-resources/data-visualisations/skills-forecast>
14. Bakhshi, H., Downing, J.M., Osborne, M.A., Schneider, P.: *The future of skills: employment in 2030*. Pearson (2017)
15. Szabó, I., Neusch, G.: Dynamic skill gap analysis using ontology matching. In: Kő, A., Francesconi, E. (eds.) *EGOVIS 2015. LNCS*, vol. 9265, pp. 231–242. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-22389-6\\_17](https://doi.org/10.1007/978-3-319-22389-6_17)
16. Ternai, K., Fodor, S., Szabó, I.: Business process matching analytics. In: Doucek, P., Basl, J., Tjoa, A.M., Raffai, M., Pavlicek, A., Detter, K. (eds.) *CONFENIS 2019. LNBIP*, vol. 375, pp. 85–94. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-37632-1\\_8](https://doi.org/10.1007/978-3-030-37632-1_8)
17. Jackson, D.A., Somers, K.M., Harvey, H.H.: Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence? *Am. Nat.* **133**, 436–453 (1989)