

Analysis of Correlating Competencies Required on the Romanian and European Union IT Workforce Markets

Cristian Georgescu

cristian.georgescu@ugal.ro

„Dunarea de Jos” University of Galati,

Considering the existence of the European Union's integrated workforce market, but also for obtaining a better interpretation of our results, in this paper we analyse correlating competencies required on the Romanian and European Union IT workforce markets. We are using data mining processes based on an association algorithm, similar to those used for shopping carts. This analysis leads us to a better allocation of our training efforts and generates a more accurate competency development plan.

Keywords: workforce market, IT competencies, data mining, association algorithm

JEL codes: M510, J240, R230

1. Introduction

In (Georgescu & Sbughea, 2012) we developed only a statistical analyse on the requirements of the IT labor market. In (Georgescu, 2014), a superior analysis has been made using data mining algorithms. The data for this analysis consists of approximately 1,200 job offers, from which over 130 competencies, which were required by employers, have been synthesized resulting in over 5,000 "job post – competency" combinations. Microsoft Association Algorithm (Microsoft, 2013), used in that paper is an implementation of the Apriori algorithm (Agrawal & Srikant, 1994), the most renowned algorithm for discovering association rules.

This algorithm was developed for shopping cart analysis aiming to discover certain patterns in customers' behavior. From the first version of this algorithm, there were hundreds of tryouts aiming to develop it and make it applicable for a wider array of situations these contributions being systemized in the paper (Han, Cheng, & Xin, 2007). An improvement of the algorithm's performances was realized by adding the item-set's weight parameter as a supplementary step in identifying them (Lei, 2012).

By carefully analyzing this algorithm we can draw a correspondence between the present issue of this paper and the standard applicability of this algorithm. Thus, competencies have been assimilated with products chosen by the buyer and the shopping carts with job posts. In this context $I = \{i_1, i_2, \dots, i_m\}$ is the set of i competencies required by the employer and $D = \{t_1, t_2, \dots, t_n\}$ represents the set of t job posts. Rephrasing our problem into shopping cart terms, employers have certain tendencies towards consumption needing specific sets of competencies that they materialize by creating a vacancy and posting it along with the competencies required for that job. It's like going to the market, picking a set of competencies and putting them into their basket.

In (Georgescu, 2015) we ran an exploratory clustering algorithm on job posts from Romania. We used the EM clustering algorithm (Dempster, Laird, & Rdin, 1977), where each job post is characterized by a number of competencies required by the employer.

Considering the existence of the European Union's integrated workforce market, but also for obtaining a better interpretation of our results, in this paper we analyse correlating competencies required on the Romanian and European Union IT workforce markets.

2. Analysis of correlating competencies required on the Romanian workforce market

In (Georgescu, 2014) we obtained, for romanian IT workforce market, 184 rules on which Figure 1 is built.

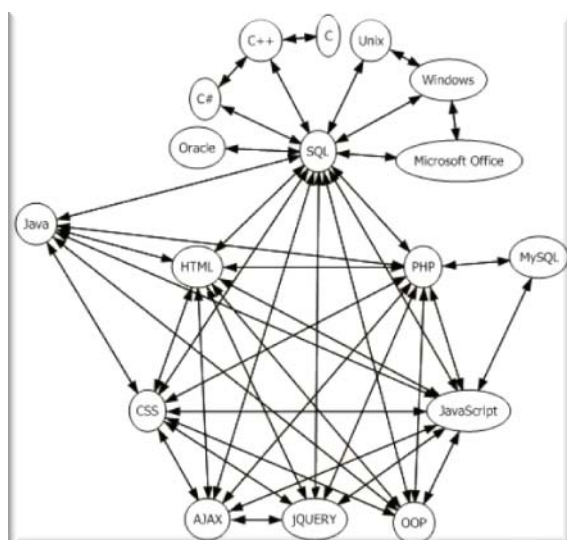


Figure 1: Dependency Network for romanian IT workforce market

A close analysis of Figure 1 and the set of associated rules helps identify three main groups of rules with common characteristics: the group of rules with negative importance, the group of rules with high importance which do not include main competencies (PHP, HTML, JavaScript, CSS and SQL) and those in close connection with them and, finally, the group of rules with the main competencies and those depending on them. These groups will be analyzed separately by filtering input data, obtained rules and by presenting output data in an appropriate manner so as to extract essential information.

The rules from the first analyzed group, the one with negative importance parameter, are presented in Figure 2. The negative value of importance represents an inverse correlation. For example, being given the SQL competency, with a probability of 14.4% jQUERY competency will not be required and the other way around, given the jQUERY competency, with a probability of 42.4% SQL will not be required.

Based on the data from Figure 2, of rules containing the SQL competency either as antecedent or as consequence, we highlight what competencies one may not have, with their respective probabilities, in case of having the SQL competency.

Analyzing, for example, the $\{C++ \rightarrow SQL\}$ rule we see that C++⁹⁹ does not help you find a job in which SQL is required, with a probability of 34%. The dependencies generated by the algorithm are not transitive. In this respect, considering the C¹⁰⁰ competency that has no direct link to SQL (Figure 1) but is strongly linked, in both ways, with C++, just as we stated that C++ competency is no help, neither will the C competency help you find a job in which SQL is required. Regardless, not the same thing can be stated about C#¹⁰¹ and Unix¹⁰² which, despite appearing next to competencies that are negatively correlated with SQL (Windows and C++), they do have a separate, positive correlation with SQL.

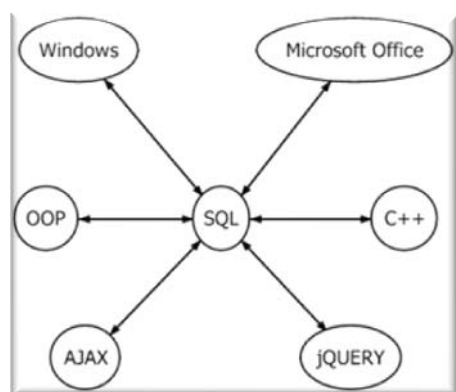


Figure 2: Competencies negatively correlated and rules with negative importance

Probability	Importance	Rules
12.40%	-0.18644316	SQL -> C++
16.50%	-0.137469741	SQL -> Windows
34.00%	-0.132310398	C++ -> SQL
36.60%	-0.100059643	Windows -> SQL
11.70%	-0.066888667	SQL -> Microsoft Office
12.40%	-0.05946068	SQL -> AJAX
19.60%	-0.059036634	SQL -> OOP
40.50%	-0.043895491	Microsoft Office -> SQL
41.00%	-0.042529222	OOP -> SQL
40.90%	-0.039168484	AJAX -> SQL
14.40%	-0.033346073	SQL -> jQUERY
42.40%	-0.022158728	jQUERY -> SQL

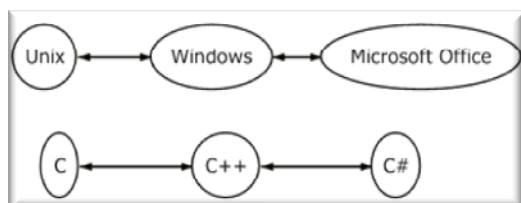
⁹⁹ C++ - object oriented programming language

¹⁰⁰ C - medium level programming language

¹⁰¹ C# - object oriented programming language, development instrument for the .NET platform

¹⁰² Unix - open-source operating system

Selecting the second group of rules, the one that does not contain the main competencies and those associated with them, we obtain Figure 3.



Probability	Importance	Rules
51.90%	1.309787835	C++ -> C
80.90%	0.957201938	C -> C++
39.70%	0.802896033	Windows -> Microsoft Office
61.90%	0.644842053	Microsoft Office -> Windows
27.50%	0.531867372	Windows -> Unix
31.10%	0.46105551	C++ -> C#
46.20%	0.445519068	Unix -> Windows
35.90%	0.441167792	C# -> C++

Figure 3: Competencies and rules not correlated with main competencies

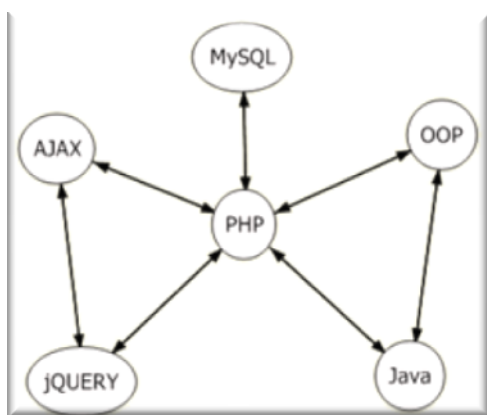
Because the following two competency groups Unix, Windows, Microsoft Office¹⁰³ and C, C++, C# are sustained by rules that, through their importance, put them in the first half of the table with rules associated to Figure 1, they will appear, within this correlation configuration, in all analyses made separately for each of the main competencies. The lack of dependencies between competencies from Figure 3 and the main competencies allow us to look at them as general knowledge requirement, which does not directly relate to the required profile for the main competencies analyzed. Therefore, by removing them from rule tables we will only get significant rules for each of the main competencies.

The last group of rules analyzed contains main competencies that will be subjected to a separate analysis in order to identify and highlight correlations and competencies associated to each of them. Using the same running parameters, in filtering conditions, when the competency is treated separately does not bring extra information, but only extracts information regarding the selected competency from the general context. Extra correlations can be obtained by modifying input parameters thus allowing for more dependencies to spur out, but this may hide other, truly interesting aspects, due to the big number of total dependencies.

Determining the cloud of competencies that are correlated with each of the main competencies only means highlighting certain correlations between competencies, obtained by filtering only, these being a subset of all correlations from Figure 1. The probability and importance of each of the 184 rules attached to Figure 1 will thus be presented separately for a better analysis.

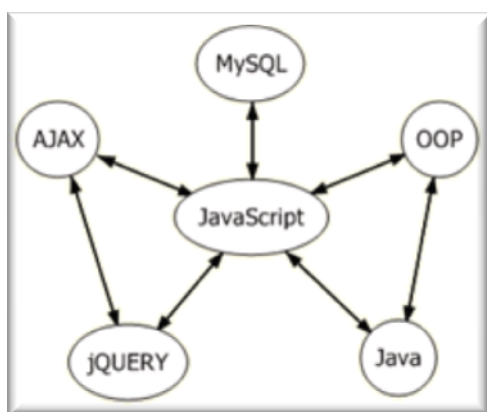
For each main competency analyzed, dependencies to the other main competencies have been removed because they were already analyzed and the purpose is to observe which other competencies contribute to forming the demand for the given competency. We have also eliminated negative correlations and rules pertaining to the group of isolated competencies analyzed before.

For each of the main competencies, there is a graphical representation of their dependencies and a table representation of the probability and importance of association rules.



Probability	Importance	Rules
15.90%	0.715377928	PHP -> MySQL
35.40%	0.566221566	jQUERY -> AJAX
39.80%	0.544258901	AJAX -> jQuery
28.40%	0.49539444	PHP -> jQuery
23.60%	0.426937059	PHP -> AJAX
71.70%	0.390766222	MySQL -> PHP
59.60%	0.344372816	jQUERY -> PHP
32.20%	0.299213099	PHP -> OOP
55.70%	0.295719779	AJAX -> PHP
26.60%	0.293635895	OOP -> Java
34.60%	0.269882563	Java -> OOP
48.20%	0.245331215	OOP -> PHP
18.80%	0.093237947	PHP -> Java
36.40%	0.074547262	Java -> PHP

¹⁰³ Microsoft Office – office software package from Microsoft

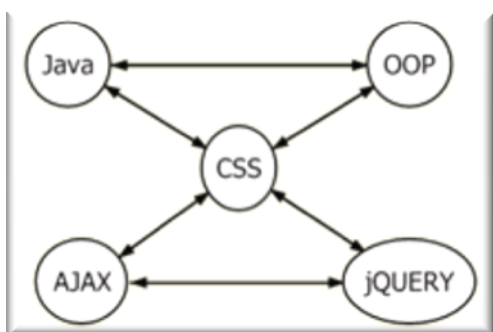


Probability	Importance	Rules
13.30%	0.621649359	JavaScript -> MySQL
35.40%	0.566221566	jQuery -> AJAX
27.30%	0.565661761	JavaScript -> jQuery
39.80%	0.544258901	AJAX -> jQuery
23.80%	0.536758325	JavaScript -> AJAX
70.70%	0.321729567	jQuery -> JavaScript
30.90%	0.309284819	JavaScript -> OOP
69.30%	0.301005375	AJAX -> JavaScript
73.90%	0.300563154	MySQL -> JavaScript
26.60%	0.293635895	OOP -> Java
34.60%	0.269882563	Java -> OOP
56.80%	0.217780632	OOP -> JavaScript
18.80%	0.103569496	JavaScript -> Java
44.90%	0.072985985	Java -> JavaScript

Figure 4: Dependencies graphs and attached tables of rules for similar competencies PHP and JavaScript

According to Figure 4, specific competencies required by PHP can be completed with rules having a complex antecedent that do not appear in the attached table due to the filtration of the other main competencies. Rules like {OOP, HTML -> PHP} with a probability of 69.1% and an importance of 0.34 or {Java, PHP -> HTML} with a probability of 84.6% and an importance of 0.37 show that PHP requires and is required by other competencies within its cloud not only directly, but also in conjunction with other main competencies.

JavaScript is no exception from the latter observation regarding the common competency group MySQL, Java, jQuery, OOP and AJAX that predict it and are predicted by it. Also it is not an exception from strengthening mutual dependencies with the main competencies marked by rules of pairs of competencies that predict it. The JavaScript competency is similar to PHP through the competencies that require it and those it requires. From Figure 4, we observe the position, in associated tables, of the rules is almost the same, having close values for importance and probability.



Probabi-lity	Importance	Rules
33.50%	0.679001104	CSS -> jQuery
35.40%	0.566221566	jQuery -> AJAX
27.10%	0.562011102	CSS -> AJAX
39.80%	0.544258901	AJAX -> jQuery
68.70%	0.447623008	jQuery -> CSS
62.50%	0.375871516	AJAX -> CSS
34.00%	0.339141683	CSS -> OOP
26.60%	0.293635895	OOP -> Java
49.60%	0.279874919	OOP -> CSS
34.60%	0.269882563	Java -> OOP
17.70%	0.056171219	CSS -> Java
33.60%	0.045828634	Java -> CSS

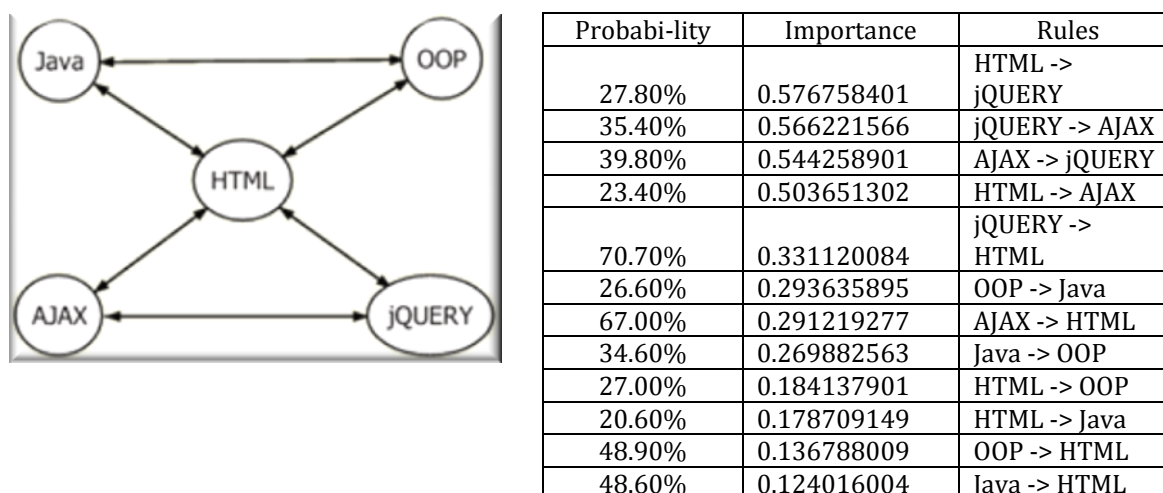


Figure 5: Dependencies graphs and tables of rules attached to similar competencies CSS and HTML

The competencies linked with CSS are the same as those for HTML, namely: jQuery, OOP, Java and AJAX. In Figure 5, we observe that rules with the same consequences of the two competencies have close values for importance and probability, being situated on similar positions in their respective tables. These explain interdependent links between main competencies HTML and CSS and links between them and other competencies, from the following level of detailed analysis, indicating value creation potential on the workforce market, when used in conjunction.

If we observe that also competencies PHP and JavaScript have just MySQL in addition to the same set of competencies, then we have a correct image for this competency group that responds to website development requirements on the workforce market.

We observe that all preceding analyses have been made in detail, taking each main competency both as input (as antecedent in a rule), positioning it on the left side of the rule, and as output (consequence in a rule), positioning it on the right side of the rule. This approach allowed the separate analysis of both the required dependencies of a certain competency and also of those that require the analyzed competency.

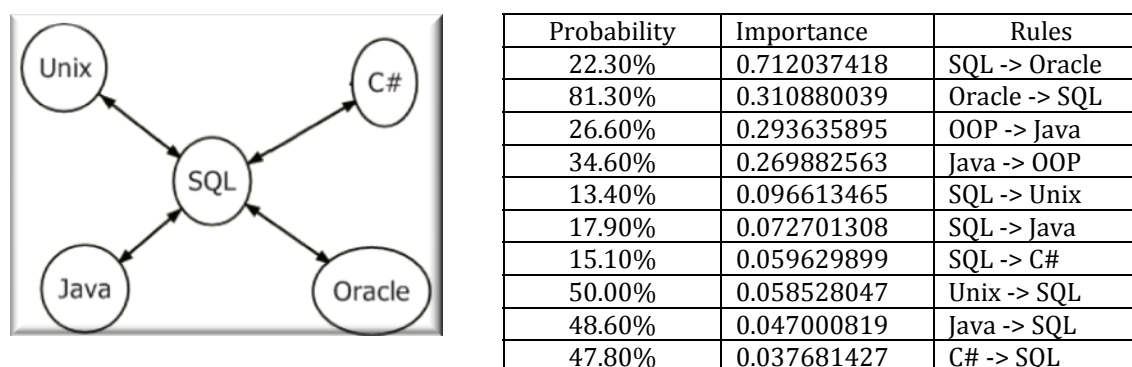


Figure 6: Dependencies graph and rules for SQL competency

The cloud of competencies that require and are required by SQL is divided into two categories: the ones negatively correlated and the ones positively correlated to SQL. Competencies with a negative correlation have already been analyzed in Figure 2. From all competencies with a positive correlation, only Java, from the second level competencies, is found linked to the main ones. Moreover, competencies like AJAX, jQuery and OOP, which are among the competencies being mutually dependent to the other main competencies (figures 4 and 5), are in the negative correlation category when it comes to SQL (Figure 2), meaning they do not require SQL with a probability of over 40%. This situation may lead us to creating a differentiated strategy when gaining competencies like SQL, on one hand, and PHP, CSS, HTML and JavaScript, on the other hand.

Regarding the isolated competency group formed by Unix, Windows, Microsoft Office and C, C++, C#, observed and commented upon beforehand (Figure 3), we now see they are dependent to SQL, by the positive correlations {SQL→C#} and {SQL→Unix}, but the other competencies: C, C++, Windows and Microsoft Office are in negative correlations with SQL. This latter find implies rethinking their position as a requirement for

general education. These competencies are a general knowledge requirement for all main competencies, but only partially for SQL: only through C# and Unix competencies.

Considering the sole existence of Oracle's¹⁰⁴ DBMS among the competencies that require and are required by SQL can be explained by an ambiguity in formulating employers' requirements or by an erroneous processing of initial data, where SQL Server's DBMS was assimilated to the simple knowledge of SQL language.

3. Analyzing requested competencies on the European Union workforce market

For this comparative analysis of competencies requested on the Romanian workforce market and those required in the European Union, we have selected a group of data that contains only job offers in IT from countries within the European Union, except Romania. The number of sets entering this analysis and the minimum probability as a rule to enter these calculations has been kept the same as in the previous analysis, of the Romanian market. Figure 7 was obtained, showing competencies required on the EU workforce market, excluding Romania, corresponding to Figure 1.

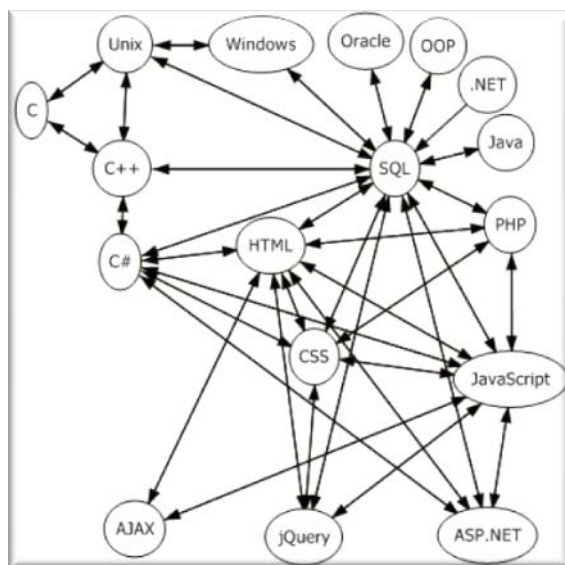


Figure 7: Dependency Network in the EU

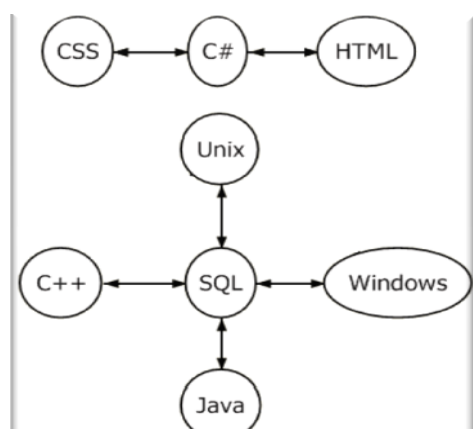
At a first look over the two diagrams, among other comments, we especially find that in Figure 7 all requirements for MySQL and Microsoft Office competencies have disappeared, being replaced with ASP.NET¹⁰⁵. Also, the OOP competency only appears positively correlated with competency SQL, without requiring or being required by the other competencies in the EU, while on the Romanian market it was negatively correlated with SQL. The position of Java language is taken over by C# which is in link, as is on the Romanian market, with C++ and C.

The same steps were followed as in the analysis of data from Romania: highlighting negative correlations between competencies, identifying groups of repeatable competencies, without being linked to the main competencies (C#, HTML, JavaScript, CSS and SQL) and the detailed analysis of the main competencies, together with those to which they are linked.

As regards the negative correlations, these change, offering another image on competencies that do not necessarily have to coexist.

¹⁰⁴ Oracle – database management system

¹⁰⁵ ASP.NET – Microsoft technology for creating web applications



Probability	Importance	Rules
29.00%	-0.014015444	CSS, JavaScript -> C#
28.60%	-0.029748704	HTML -> C#
27.20%	-0.030341139	C# -> HTML
27.90%	-0.036628895	HTML, JavaScript -> C#
26.30%	-0.038396391	C#, SQL -> HTML
27.20%	-0.049338789	CSS, HTML -> C#
47.80%	-0.058203558	Unix -> SQL
26.70%	-0.058594715	CSS -> C#
18.40%	-0.065879464	C# -> CSS
46.80%	-0.067857187	Java -> SQL
44.30%	-0.094210236	Windows -> SQL
14.90%	-0.099169095	SQL -> Unix
13.10%	-0.116803951	SQL -> Java
12.20%	-0.159356301	SQL -> Windows
38.60%	-0.170762426	C++ -> SQL
15.30%	-0.258740934	SQL -> C++

Figure 8: Negatively correlated competencies and rules with negative importance in the EU

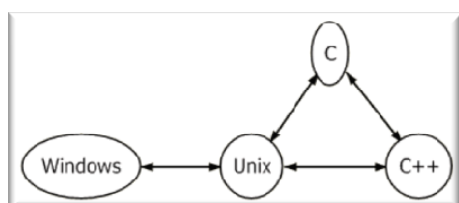
Note that in Figure 8, apart from direct dependencies represented by rules with one antecedent and one consequence, there is a combination of two competencies on the left side of the rule, which also represents a negative correlation and it links the two subgroups.

C#, SQL $\rightarrow$ HTML.

Another aspect that needs consideration is the negative correlation between main competencies C# and CSS, HTML, which goes to show that a requirement in C# needs specialization in this direction, without many other collateral competencies.

An observation that is equally true for the same analysis of negative correlations from Romania refers to the rules' low degree of importance and probabilities well below 50%. This shows that we shouldn't, in any case, talk about an interdiction to associate competencies from Figures 2 and 8, but rather a suggestion regarding the disassociation of certain competencies.

From Figure 7, but especially from the rules that base this graphical representation, we observe a group of competencies similar to that identified in the analysis of Romania, in Figure 3. Isolating this group, we obtain Figure 9 with competencies that can be considered fundamental competencies, mandatory for sustaining any of the main competencies. Comparing this group to the similar one obtained by analyzing data on the Romanian market, we observe that competencies Microsoft Office and C# have disappeared, the latter becoming a main competency. Competencies C/C++ and Windows/Unix maintain, strengthening the observation that on both markets, the demands regarding general education remain mostly the same.

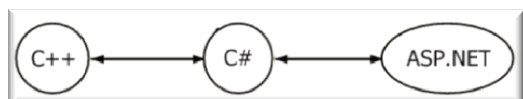


Probability	Importance	Rules
40.90%	0.953741799	C++ -> C
72.00%	0.691403596	C -> C++
30.40%	0.554372552	Unix -> C
36.20%	0.53584275	Unix -> Windows
41.00%	0.512648638	Windows -> Unix
42.00%	0.499704343	C -> Unix
29.50%	0.349542422	C++ -> Unix
37.70%	0.321094341	Unix -> C++

Figure 9: Competencies and rules uncorrelated with main competencies

The separate analysis for each of the main competencies, by hiding the others along with those associated, will be highlighted both graphically and through their associated rules. In all figures to follow, the pattern from Figure 9 along with its rules will exist, but will be removed. Thus we obtain an easier to analyze image for each competency analyzed.

For example, in Figure 10 competency C# (which replaced competency PHP from Romania) is in bidirectional dependence, with significant probabilities and importance, with C++ and ASP.NET which allows us to say that it is an extension of the fundamental education mentioned before, following the direction of Microsoft technology for creating web applications.



Probability	Importance	Rules
30.40%	0.720704954	C# -> ASP.NET
70.40%	0.457102478	ASP.NET -> C#
24.80%	0.101811253	C# -> C++
35.20%	0.089057707	C++ -> C#

Figure 10: Dependencies graph and associated rules for competency C# in the EU

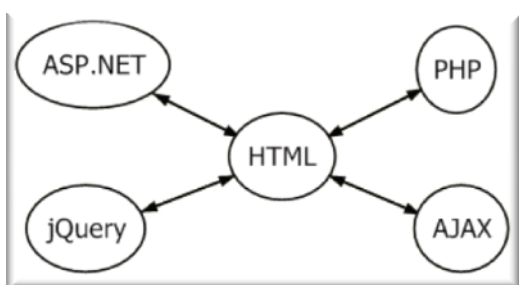
The separate analysis of competency CSS from Figure 11 shows a similar situation to that from Figure 10 in terms of precise, even restrictive phrasing of competency combination. If we also add the links with the main competencies HTML and JavaScript, through the PHP competency (Figure 12), we can conclude that CSS is a competency required for web applications development, following the Java development line.



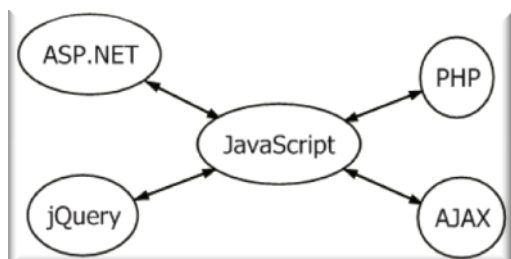
Probability	Importance	Rules
29.10%	0.989004616	CSS -> jQuery
73.50%	0.648342521	jQuery -> CSS
32.60%	0.517915326	CSS -> PHP
46.70%	0.450604295	PHP -> CSS

Figure 11: Dependencies graph and associated rules for competency CSS in the EU

Regarding competency in HTML, the situation is identical to competency JavaScript (Figure 12) and we observe main competencies HTML and JavaScript forming a subgroup due to mutual dependencies and because they require and are required by the same other competencies. CSS and C# can also be added to these competencies, the first one, as it is interdependent with the two main competencies through PHP and jQuery, while the second is interdependent with the two main competencies through ASP.NET.



22.70%	0.934039123	HTML -> jQuery
17.60%	0.88729572	HTML -> AJAX
31.90%	0.61930785	HTML -> PHP
79.40%	0.504435945	jQuery -> HTML
77.80%	0.47433909	AJAX -> HTML
63.30%	0.436309063	PHP -> HTML
20.20%	0.296549394	HTML -> ASP.NET
44.40%	0.226189319	ASP.NET -> HTML

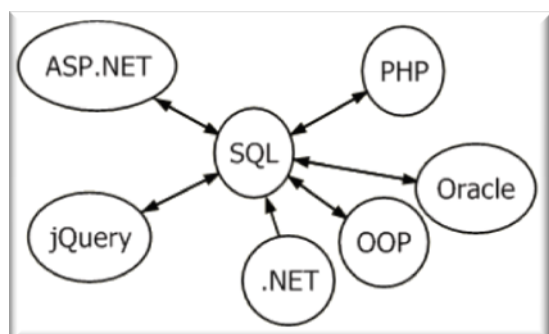


Probability	Importance	Rules
18.00%	1.026690373	JavaScript -> AJAX
21.10%	0.88951718	JavaScript -> jQuery
28.90%	0.54502149	JavaScript -> PHP
85.20%	0.48245698	AJAX -> JavaScript
79.40%	0.464318722	jQuery -> JavaScript
61.70%	0.375054078	PHP -> JavaScript
18.00%	0.220510399	JavaScript -> ASP.NET
42.60%	0.16542592	ASP.NET -> JavaScript

Figure 12: Dependencies graph and associated rules for competencies HTML and JavaScript in the EU

Competency in SQL, as was the case for the Romanian market, implies a series of negative correlations already present in Figure 8. Negative correlations are realized with the competency pattern of C++, Unix, Windows and C, as it is presented in Figure 9, to which we add the Java language but, due to the relatively low importance parameter, these rules should not be interpreted in the strict sense of exclusion.

More important though are the positive correlations. In this respect, we observe the link with competencies ASP.NET, jQuery and PHP that also exists for the other main competencies HTML, JavaScript and, indirectly, for CSS, which are dedicated to web applications. This aspect, corroborated with the disappearance of the requirement for competency MySQL, leads us to conclude that the database support for this category of applications, in EU requirements, is ensured by SQL and Oracle.



Probability	Importance	Rules
23.40%	0.43138066	SQL -> Oracle
17.10%	0.298169397	SQL -> ASP.NET
10.80%	0.294101035	SQL -> jQuery
18.00%	0.228118274	SQL -> PHP
76.50%	0.187760963	Oracle -> SQL
11.30%	0.176435801	SQL -> OOP
70.40%	0.134413422	ASP.NET -> SQL
70.60%	0.124847794	jQuery -> SQL
66.70%	0.109391075	PHP -> SQL
64.10%	0.081865651	OOP -> SQL
57.90%	0.033963587	.NET -> SQL

Figure 13: Dependencies graph and associated rules for competency SQL in the EU

On the other hand, SQL is also required by the other main competencies (HTML, JavaScript and C#) according to their respective utilization domain. We again observe a better orientation of employers, expressed through the required combination of competencies and a better indication of required competencies.

4. Conclusions

The comparative analysis under territorial aspect between Romania, considered starting point and the other countries within the EU, considered as ulterior analysis, has given us a number of conclusions.

The disappearance of Microsoft Office competency from requirements in the IT field within Europe (Figure 3), as compared to Romania (Figure 1) actually reflects a higher training level. Office knowledge is requested for organizational levels where it is actually used and not at the IT level where they are considered to be implicit.

Disappearance of MySQL competency along with the lack of PHP from the main competencies shows us that the PHP and MySQL websites era is about to set, leaving way for more complex instruments for web design.

Having replaced requirements for Java language with those for C# and positioning the latter among main competencies, in addition to the persistence of the C++, C, Unix, Windows competency pattern might tempt us to assume a tendency based on predominant software development on the C direction to the detriment of Java. Things are not as easy and clear because JavaScript is still part of the main competencies in analyses regarding both Romania and the other EU countries. We can state that closer to the truth would be eliminating the simple knowledge of Java language and orienting towards using web development technology based on JavaScript.

The reduced importance of generic competency OOP can be attributed to a specialization trend through which employers request the specific programming tool and not general knowledge of object programming. Actually OOP is the basis for the new programming paradigm considered a fundamental element in programming and, consequently, implicitly required and not explicitly, as a competence.

ASP.NET appears among the competencies required on the European market and this must be put in context with Microsoft's position and seen in the light of its competition with other companies that promote other development solutions.

A conclusion that must not be neglected is that of negative conditioning between competencies, in other words "what must not be mandatory correlated", analysis which leads us to a better allocation of training efforts and competency gaining.

Bibliography

1. Agrawal, R., & Srikant, R. (1994). *Fast Algorithms for Mining Association Rules. Proceedings of the 20th VLDB Conference*, (pp. 487-499). Santiago, Chile.
2. Dempstern, A. P., Laird, M., & Rdin, D. B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1., 1-36.
3. Georgescu, C. (2014). *Analyzing the Mix of Competencies Required for IT Labor Market, Using Data Mining Algorithms. Analyzing the Mix of Competencies Required for IT Labor Market, Using Data Mining Algorithms; Proceeding of the 13th International Conference on Informatics in Economy, 2014. Bucharest.*
4. Georgescu, C. (2015). *Identifying the competency mix required for the Romanian IT labor market. The risk in contemporary economy, FEAA Galati, 2015. Galati.*
5. Georgescu, C., & Sbughea, C. (2012). *Data warehouse for professional skills required on the IT labor market. The risk in contemporary economy. ISSN 2067-0532.*
6. Han, J., Cheng, H., & Xin, D. (2007). *Frequent pattern mining: current status and future. Data Mining and Knowledge Discovery*, 55-86.
7. Lei, C. (2012). *The Research of Data Mining Algorithm Based on Association Rules. The 2nd International Conference on Computer Application and System Modeling.*
8. Microsoft. (2013). *Microsoft Association Algorithm Technical Reference. Retrieved June 2013, from MSDN.*