

УДК 004.8

© Л.С. Коряшкіна, В.С. Чернишенко, М.О. Віноградов, А.Е. Скрипченко

## **ЗАСТОСУВАННЯ МЕТОДІВ DATA MINING ПРИ АНАЛІЗІ ВИМОГ РОБОТОДАВЦІВ ЩОДО ЯКОСТІ ПІДГОТОВКИ СПЕЦІАЛІСТІВ**

© L. Koriashkina, V. Chernyshenko, M. Vinogradov, A. Skrypchenko

### **DATA MINING METHODS IN ANALYZING EMPLOYERS REQUIREMENTS FOR THE QUALITY OF SPECIALISTS TRAINING**

Представлений один підхід щодо вивчення вимог роботодавців стосовно компетенцій випускників вищих навчальних закладів, який включає: збір і систематизацію необхідної інформації, вивчення і аналіз отриманих даних за допомогою методів Data Mining, висновки і рекомендації щодо наповнення освітніх програм підготовки відповідних фахівців, які можна отримати з результатів інтелектуального аналізу даних

Представлен один подход к изучению требований работодателей относительно компетенций выпускников высших учебных заведений, включающий: сбор и систематизацию необходимой информации, изучение и анализ полученных данных с помощью методов Data Mining, выводы и рекомендации по наполнению образовательных программ подготовки соответствующих специалистов, которые можно получить из результатов интеллектуального анализа данных

**Вступ.** У сучасних умовах однією з ключових задач модернізації європейської вищої освіти є покращення її якості та релевантності [1], що може бути забезпечено, зокрема, шляхом широкого залучення роботодавців та інституцій ринку праці до розроблення та запровадження освітніх програм, використання потенціалу сучасних інформаційно-комунікаційних технологій (ІКТ) для забезпечення індивідуалізації навчання, удосконалення методів викладання та дослідження.

Згідно Політиці забезпечення якості освіти НГУ [2], а також інших навчальних закладів, якість вищої освіти визначається наступними показниками: відповідністю змісту вищої освіти щодо вимог роботодавців і запитів здобувачів; відповідністю Національній рамці кваліфікацій; рівнем застосування компетентнісного підходу до проектування вищої освіти; визначенням й оприлюдненням очікуваних результатів навчання шукачів за всіма видами навчальної діяльності; безпосереднім зв'язком результатів навчання з компетенціями випускників, регламентованих стандартами вищої освіти та освітніми програмами за спеціальностями (спеціалізаціями); рівнем інтеграції вищої освіти з наукою і виробництвом.

Сьогодні вищі навчальні заклади мають здійснювати регулярний моніторинг освітніх програм з метою забезпечення належного рівня освітніх послуг та створення сприятливого й ефективного навчального середовища, що передбачає оцінювання таких параметрів: забезпечення актуальності освітніх програм; зміна потреб суспільства; навантаження, навчальні досягнення й успішність студентів; ефективність процедур оцінювання студентів; очікування, потреби та

задоволеність студентів освітніми програмами; навчальне середовище, служба підтримки студентів, їхня відповідність цілям програми [3].

У зв'язку з цим актуальним постає пошук ефективних способів вивчення інноваційних процесів у галузі праці, запитів роботодавців до системи вищої освіти. Динамічна зміна умов професійної діяльності вимагає від вищих навчальних закладів систематичного вивчення рівня задоволеності роботодавців як зацікавленої сторони в якісній підготовці фахівців. Важливим напрямком досліджень є розробка методів і алгоритмів щодо вирішення завдань аналізу якості вищої освіти, маркетингових досліджень працевлаштування випускників, створення механізмів, що дозволяють аналізувати, чи відповідають очікування студента щодо його компетенцій на передбачувану посаду вимогам роботодавців, враховувати думку випускників, які вже отримали цю роботу.

В даній роботі представлений один підхід щодо вивчення вимог роботодавців стосовно компетенцій випускників вищих навчальних закладів, який включає: збір і систематизацію необхідної інформації, вивчення і аналіз отриманих даних за допомогою методів Data Mining, висновки і рекомендації щодо наповнення освітніх програм підготовки відповідних фахівців, які можна отримати (винести) з результатів інтелектуального аналізу даних. Роботу виконано у рамках проекту «Система забезпечення якості в Україні: розвиток на основі ENQA стандартів та норм (QUAERE)».

**Збір і систематизація інформації.** В якості досліджуваних були обрані вакансії з ринку праці в галузі комп'ютерних інформаційних технологій і системного аналізу, а саме такі: системний аналітик, менеджер проектів, фахівець з інформаційних ресурсів (ІР), фахівець з інформаційних систем (ІС).

Вочевидь, виявлення (ідентифікація) компетенцій, що пред'являються роботодавцями до випускників навчальних вищих закладів або претендентів на ту чи іншу посаду, має здійснюватися з використанням анкетування, яке в свою чергу передбачає вирішення низки типових для соціологічних досліджень питань: організаційних, методологічних і методичних. До них, насамперед, належать розробка анкети, визначення складу респондентів (оптимального на якісному і кількісному рівнях), вибір способів опитування та збору анкет, методів обробки та аналізу інформації, а також форми представлення результатів.

З урахуванням того, що основною метою даної роботи є дослідження можливості застосування методів математичного моделювання і Data Mining для проведення аналізу вимог роботодавців, то, не претендуючи на повноту і коректність збору даних, з різних сайтів роботодавців протягом одного кварталу поточного року були вивчені дві сотні листів роботодавців із зазначенням знань, умінь і навичок, якими має володіти претендент на одну із вказаних вище посад. Кожна професія у вибірці представлена півсотньою листів-запитів. Вся зібрана інформація структурована у вигляді розробленої в MS Access бази даних «Вимоги роботодавців». Загалом база налічує 1300 записів.

Респондент в БД представлений двома атрибутами: ідентифікаційний номер і тип, який може бути закодований наступним чином: 1 – роботодавець; 2 – випускник вузу, який працює за фахом; 3 – викладач вузу; 4 – студент. Кожно-

му типу респондента можна поставити ваговий коефіцієнт, який у свою чергу може бути врахований при обчисленні тих чи інших параметрів (наприклад, при підрахунку голосів за конкретну компетенцію). Підкреслимо, що в даній роботі представлені результати досліджень за голосуванням респондентів лише першого типу.

Нагадаємо, що під вимогами роботодавців розуміються їх очікування щодо компетенцій працівників конкретної професії та конкретного посадового рівня. Сучасна система професійної освіти повинна мати чіткі уявлення про те, для яких можливих видів професійної діяльності вони здійснюють підготовку, а не просто готувати уніфікованих працівників, що володіють компетенціями, частина з яких, можливо, застаріла. У зв'язку з цим аналіз потреб у знаннях і уміннях необхідно проводити на максимально можливій вибірці підприємств, орієнтованих на розвиток, оскільки тільки так можна сформулювати орієнтири на кращі зразки професійної діяльності. Саме тому при зборі даних враховувалася також інформація про те, до якої галузі народного господарства (промисловості, виробництва, науки, знання) належить роботодавець. Перелік галузей народного господарства, які включено в БД, наведений у табл. 1.

Важливість цієї позиція також обумовлена можливістю враховувати під час підрахунку голосів про актуальність будь-якої компетенції для конкретної професії факт безпосереднього чи опосередкованого відношення до даної професії респондента, який виносить судження. Його оцінка у першому випадку може «посилюватися», у другому – «послаблюватися».

Таблиця 1

Галузі народного господарства, які включено в БД «Вимоги роботодавців»

Код_галузі	Найменування_галузі	Код_галузі	Найменування_галузі
1	Житлово-комунальне господарство	21	Медіа
2	Невиробничі види побутового обслуговування населення	22	Аналітика
3	Охорона здоров'я, фізична культура і соціальне забезпечення	23	Електроенергетика
4	Освіта	24	Паливна промисловість
5	Фінанси, кредит, страхування, пенсійне забезпечення	25	Чорна металургія
6	Культура і мистецтво	26	Кольорова металургія
7	Наука і наукове обслуговування	27	Машинобудування і металообробка
8	Управління	28	Хімічна та нафтохімічна промисловості
9	Партійні та громадські об'єднання	29	Лісова, деревообробна і целюлозно-паперова промисловість
10	Геологорозвідка	30	Промисловість будівельних матеріалів
11	Торгівля	31	Промисловість будівельних конструкцій і деталей
12	Будівництво	32	Скляна та фарфоро-фаянсова промисловості
13	Транспорт і логістика	33	Легка промисловість
14	Сервісні компанії	34	Харчова промисловість
15	Телекомунікації	35	Мікробіологічна промисловість
16	ІТ	36	Мукомельно-круп'яна і комбікормова промисловості
17	Дизайн	37	Медична промисловість (виробництво медичної апаратури та інструментів)
18	Реклама	38	Фармацевтика
19	Охорона	39	Інші галузі промисловості
20	Туризм		

Структурними одиницями у базі даних є також професія, про компетенції якої виноситься судження (назва професії наводиться відповідно до Класифікатора професій), і компетенції, представлені списком знань і умінь (вид компетенції). Самі компетенції розділені за такими категоріями: професійно-нормативні, професійно-виробничі, управлінські, соціальні, особистісні [4, 5]. З метою використання цих даних у подальшому аналізі зібраної інформації за допомогою методів Data Mining зручно кожній компетенції поставити у відповідність деякий шифр, наприклад, як наведено на рис. 1. Шифри компетенцій у розробленій БД включені до окремої таблиці.

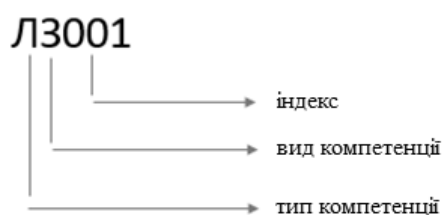


Рис. 1. Склад шифру компетенції

**Зауваження 1** (про відмінності між професійно-нормативними та професійно-виробничими компетенціями). Тут передбачається, що перші – це ті знання і вміння, які визначені у професійному стандарті для кожної з галузей, другі – це знання і вміння, якими повинен володіти фахівець з точки зору роботодавця. Останні беруться з оголошень вакансій на ринку праці із зазначенням вимог до претендентів.

БД «Вимоги роботодавців» має насамперед інформаційний характер. Завдяки цьому за допомогою створення різних запитів можна легко і швидко отримати необхідну інформацію для подальшої її обробки, наприклад, за допомогою аналітичної платформи Deductor.

Як відомо, існує три класичних схеми побудови баз даних для аналізу: «зірка», «сузір'я», «сніжинка». Схема створеної БД – «Зірка». Для цієї схеми характерна єдина велика центральна таблиця фактів і по одній таблиці для кожного вимірювання.

**Математичні основи аналізу вимог роботодавців щодо компетенцій за різними профілями підготовки спеціалістів.** Залежно від того, як саме відображається інформація про голосування за ті чи інші компетенції, по-різному можна сформулювати математичну постановку задачі пошуку переліків найбільш важливих (актуальних) компетенцій [5]. В одному випадку це може бути система обмежень у формі рівностей або нерівностей. Другий варіант передбачає формулювання задачі в термінах теорії множин.

Представимо ці два підходи.

I. Введемо наступні позначення:

$K$  – кількість відповідей респондентів; з довільною такою відповіддю будемо пов'язувати індекс  $k, k = \overline{1, K}$ ;

$N$  – кількість типів можливих респондентів;

$Q$  – кількість галузей народного господарства, виробництва або науки, що розглядається; кожній галузі відповідає свій індекс  $q, q = \overline{1, Q}$ ;

$P^q$  – кількість професій, асоційованих з галузю  $q, q = \overline{1, Q}$ ; з кожною професією з галузі  $q, q = \overline{1, Q}$  будемо пов'язувати індекс  $p^q, p^q = \overline{1, P^q}$ ;

$L$  – число типів компетенцій для кожної професії; (тут  $L = 5$ ;  $l = 1$  відповідає професійно-нормативним компетенціям,  $l = 2$  – професійно-виробничим і т.д.);

$S_{p^q}^l, U_{p^q}^l$  – кількість видів знань, умінь  $l$ -ого типу відповідно, якими має володіти претендент на посаду  $p^q$ ; самі знання і уміння будемо позначати відповідно  $Lore_{p^q, s}^l, s = \overline{1, S_{p^q}^l}, Skills_{p^q, u}^l, u = \overline{1, U_{p^q}^l}$

$Type\_R = \{t\_rsp_1, t\_rsp_2, \dots, t\_rsp_N\}$  – множина всіх типів респондентів;

$Branch = \{br_1, br_2, \dots, br_Q\}$  – множина усіх галузей промисловості, науки, виробництва, що розглядаються;

$Jobs^{br_q} = \{job_1^{br_q}, job_2^{br_q}, \dots, job_{p^q}^{br_q}\}, q = \overline{1, Q}$  – множина професій, асоційованих з  $q$ -ою галуззю.

Голосування  $k$ -го респондента за набір компетенцій  $l$ -ого типу (а саме, знань та умінь) за обраною професією будемо представляти векторами відповідних розмірностей, які складаються з одиниць і нулів залежно від того, віддав чи ні свій голос респондент за відповідну компетенцію:

$a_{p^q}^{k,l} = \{a_{p^q,1}^{k,l}, a_{p^q,2}^{k,l}, \dots, a_{p^q,s}^{k,l}\}, S = S_{p^q}^l$  – вектор голосів за знання  $l$ -ого типу, якими має володіти претендент на посаду  $p^q$ ;  $a_s^{k,l} \in \{0,1\} \forall s = \overline{1, S_{p^q}^l}$ ;

$b_{p^q}^{k,l} = \{b_{p^q,1}^{k,l}, b_{p^q,2}^{k,l}, \dots, b_{p^q,u}^{k,l}\}, U = U_{p^q}^l$  – вектор голосів за уміння  $l$ -ого типу, якими має володіти претендент на посаду  $p^q$ ;  $b_u^{k,l} \in \{0,1\} \forall u = \overline{1, U_{p^q}^l}$ .

Зафіксуємо довільну професію  $job_q^{br_j}$ , голосування за компетенції якої було проведено. Їй відповідає галузь  $br_j \in Branch$  з порядковим номером  $q, (br_j = br_q, q = \overline{1, P^q})$ , а також порядковий номер  $p^q$  в наборі професій з цієї галузі.

Розглянемо питання про те, як можна оцінити ту чи іншу компетенцію обраної професії за результатами опитування.

Введемо до розгляду вагові коефіцієнти респондентів двох видів:

1)  $w_k \in \{W_1, W_2, \dots, W_N\} \forall k = \overline{1, K}$ , де  $W_n$  – «вага» респондента  $n$ -го типу,  $n = \overline{1, N}$ ; тобто  $w_k = W_n$ , якщо тип респондента дорівнює  $t\_rsp_n$ ;

2)  $g_k \in \{G_{enhance}, G_{lower}\} \forall k = \overline{1, K}$  – множник, що посилює або зменшує голос респондента при визначенні загальної оцінки компетенції професії, а саме:



$$g_k = \begin{cases} G_{enhance}, & \text{якщо } br^k = br_j, \\ G_{lower}, & \text{якщо } br^k \neq br_j, \end{cases} \quad \forall k = \overline{1, K}.$$

Слід зауважити, що допустимі значення введених вагових коефіцієнтів є параметрами алгоритму оцінювання і вибору найактуальніших компетенцій професій.

З урахуванням усіх введених позначень і величин для кожної компетенції будемо розраховувати оцінку знань і умінь  $l$ -ого типу за наступними формулами:

$$Rating_{p^q, s}^l = \sum_{k: job^k = job_q^{br-j}} w_k g_k a_{p^q, s}^{k, l}, \quad Rating_{p^q, s}^l = \sum_{k: job^k = job_q^{br-j}} w_k g_k b_{p^q, s}^{k, l}.$$

Далі можливі декілька варіантів відбору важливих (актуальних) компетенцій.

Перший варіант передбачає виключення тих знань і умінь, чиї оцінки  $Rating_{p^q, s}^l$  виявились меншими за деяке заздалегідь визначене порогове значення  $\delta > 0$ . При цьому вказаний поріг можна задавати залежно від кількості респондентів, які проголосували за відповідну професію. Отже, набір найважливіших (актуальних) компетенцій для професії  $p^q, p^q = \overline{1, P^q}$ , може бути визначений у такий спосіб:

$$\begin{aligned} Lore_{p^q}^l &= \left\{ Lore_{p^q, s}^l : Rating_{p^q, s}^l > \delta, s = \overline{1, S_{p^q}^l} \right\}, \\ Skills_{p^q}^l &= \left\{ Skills_{p^q, u}^l : Rating_{p^q, s}^l > \delta, u = \overline{1, U_{p^q}^l} \right\}. \end{aligned} \quad (1)$$

За другим варіантом відбору актуальних компетенцій здійснюється вибір із всього представленого переліку знань і умінь  $l$ -ого типу для професії із порядковим номером  $p^q$  лише тих  $D_q^l$  компонент (позначимо набір їхніх номерів через  $\sigma = \{s_1, s_2, \dots, s_{D_q^l}\}$ ), для яких виконуються наступні нерівності:

$$Rating_{p^q, s}^l \leq Rating_{p^q, r}^l \quad \forall s \in \sigma, r \in \{1, 2, \dots, S_{p^q}^l\} \setminus \sigma. \quad (2)$$

Звичайно, можна при формуванні раціонального (найбільш затребуваного) переліку компетенцій з певної професії обмежитися лише підрахунком голосів за ту чи іншу компетенцію з урахуванням вищезгаданих вагових коефіцієнтів. Однак виявляється корисним отримання з результатів голосувань цілої групи компетенцій, що часто зустрічаються, особливо, якщо ці групи були відзначені респондентами, які мають безпосереднє відношення до даної професії. Це дозволить відкинути ті поодинокі компетенції, за які проголосувало досить багато респондентів, але ці респонденти не пов'язані з даною професією і мають або «стороннє», або недостатнє (проголосував студент) уявлення про неї.

Крім того, можливий варіант, що за певну компетенцію голосують конкретні типи респондентів. Якщо це – роботодавці або представники ринку професії, то дану компетенцію варто розглядати як важливу і враховувати її при розробці або корегуванні освітньої програми. Якщо ж за компетенцію голосують лише представники освіти або студенти, то слід рекомендувати укладачам кваліфікаційних рамок професій переглянути необхідність включення до них такої компетенції.

Для виявлення вище описані ситуації потрібно: 1) для кожної професії знайти набори компетенцій, що найчастіше зустрічаються; 2) сформувавши базу асоціативних правил і оцінити силу зв'язку між ознаками; 3) побудувати дерево рішень, за допомогою якого можна простежити зворотний зв'язок між наборами компетенцій, що часто зустрічаються, і категоріями (типом і галуззю) тих респондентів, хто проголосував за ці набори; 4) побудувати дерево рішень, за допомогою якого можна виявити взаємозв'язок між професіями за однаковими наборами компетенцій, що їм властиві, особливо з точки зору роботодавців.

Корисним також виявляється проаналізувати зв'язок між галуззю, до якої відноситься респондент, і професією, про яку він виносить своє судження. Якщо галузь професії і галузь респондента не збігаються, і при цьому таких респондентів досить велика кількість, то можна зробити висновок про те, що дана професія затребувана у відповідній галузі, і потрібно цей факт відобразити при складанні програми підготовки фахівців з даного напрямку. Можливо, слід внести якісь корективи, як до назви професії, так і до їхнього змісту, щоб уникнути дублювання або плутанини при класифікації професій.

Перейдемо до розгляду другого підходу до аналізу результатів опитування осіб, зацікавлених в раціональному переліку компетенцій, притаманних тим чи іншим професіям, заснованого на формулюванні і розв'язанні задачі пошуку наборів, що часто зустрічаються у великій базі даних, і складанні асоціативних правил.

II. Сформулюємо математичну модель задачі, використовуючи введені раніше позначення, а також апарат теорії множин. Також наведемо деякі формальні визначення.

Нехай  $I = \{i_1, i_2, \dots, i_M\}$  – множина усіх можливих компетенцій, що включені у конкретну базу даних. В теорії інтелектуального аналізу даних таку сукупність прийнято називати множиною сутностей. Сюди будемо включати усі знання і уміння, які притаманні хоча б одній професії, про компетенції якої виносяться судження. Крім того, сюди ж включимо і перелік усіх розглянутих професій.

I, отже, кількість елементів множини сутностей дорівнює

$$M = \sum_{q=1}^Q \left( \sum_{p^q=1}^{P^q} \sum_{l=1}^5 (S_{p^q}^l + U_{p^q}^l) + P^q \right).$$

Якщо завдання полягає в знаходженні найбільш затребуваних наборів знань і умінь, характерних саме для конкретної професії, то можна формувати множини компетенцій для кожної розглянутої професії окремо і позначати їх у такий

спосіб:  $I^{P^q} = \{i_1^{P^q}, i_2^{P^q}, \dots, i_{M^{P^q}}^{P^q}\}$ ,  $P^q = \overline{1, P^q}$ , де  $M^{P^q} = \sum_{l=1}^5 (S_{p^q}^l + U_{p^q}^l)$ .

Базу даних будемо розглядати як множину транзакцій  $D = \{d_1, d_2, \dots, d_K\}$ , кожна з яких складається з множини елементів  $I \subset I$  та унікального ідентифікатора  $tid$ . Множина усіх таких ідентифікаторів  $TID = \{tid_1, \dots, tid_K\}$ .

При складанні ідентифікатора транзакції, який, взагалі кажучи, є певним кодом респондента, можна враховувати тип респондента, галузь науки або галузі віт-

чизняного виробництва, до якої він відноситься, а також порядковий номер у голосуванні.

Транзакція  $d = (tid, I)$  містить множину  $X$ , якщо  $X \subseteq I$ .

**Покриття** множини  $X$  у базі даних  $D$  складається з множини транзакцій, які містять  $X$ :  $cover(X, D) = \{tid \mid (tid, I) \in D, X \subseteq I\}$ .

**Підтримкою множини**  $X$  у базі даних  $D$  називається кількість транзакцій у покритті множини  $X$  у базі  $D$ :  $support(X, D) = |cover(X, D)|$ .

**Частотою множини**  $X$  у базі даних  $D$  називається ймовірність, з якої множина  $X$  зустрічається в транзакціях  $d \in D$ :

$$frequency(X, D) = P(X) = \frac{support(X, D)}{|D|}.$$

Задача пошуку наборів сутностей (компетенцій), що найчастіше зустрічаються, формулюється у такий спосіб. Нехай задана множина сутностей  $I$ , є база даних  $D$  на множині  $I$ , задано значення мінімальної підтримки наборів  $Supp_{min}$ .

Потрібно знайти множину сутностей, для якої

$$F(D, Supp_{min}) = \{X \subseteq I \mid support(X, D) \geq Supp_{min}\}.$$

Асоціативне правило – це вираз вигляду  $X \Rightarrow Z$ , де  $X$  та  $Z$  – підмножина з множини  $I$ , до того ж  $X \cap Z = \{\}$ .

Підтримка асоціативного правила – це підтримка множини  $X \cup Z$  в  $D$ .

Достовірністю асоціативного правила  $X \Rightarrow Z$  у базі даних  $D$  називається умовна ймовірність того, що за умови підтримки даною транзакцією множини  $X$ , множина  $Z$  також підтримується цією транзакцією:

$$confidence(X \Rightarrow Z, D) = P(Z|X) = \frac{support(X \cup Z, D)}{support(X, D)}.$$

Правило будемо називати достовірним, якщо його достовірність перевищує мінімально заданий поріг  $\gamma$  (ще один параметр методу, що задається дослідником).

Задача пошуку асоціативних правил в базі даних  $D$  формулюється так. Нехай задана множина суутностей (компетенцій)  $I$ , є база даних  $D$  на множині  $I$ , задані мінімальні значення підтримки шуканих наборів, і достовірності  $Supp_{min}$  і  $\gamma$ . Потрібно знайти набір правил

$$(R)(D, Supp_{min}, \gamma) = \{X \Rightarrow Z \mid X \cap Z = \{\}, X \cup Z \in F(X, D), confidence(X \Rightarrow Z, D) \geq \gamma\}.$$

У термінах даної задачі пошуку асоціативних правил при аналізі даних про голосування респондентів, достовірність правила – це ймовірність того, що респондент проголосує за набір компетенцій  $Z$ , якщо при цьому він голосує за набір компетенцій  $X$ .

Таким чином, проблема знаходження асоціативних правил містить в собі дві підзадачі:

1. Знайти всі комбінації наборів, чия підтримка більше, ніж задана.



2. На основі отриманих наборів, з огляду на задану мінімальну достовірність правила, скласти кінцевий набір асоціативних правил.

Знаходження всіх наборів з підтримкою, вище заданої, через кількість таких наборів, є вельми нетривіальним. Число всіх можливих наборів  $2^M$ .

З обчислювальної точки зору, друга підзадача повністю залежить від першої, тому в більшості наукових публікацій у цій галузі намагаються знайти оптимальний (в певному сенсі) розв'язок першої задачі. Однак слід зазначити, що і друга підзадача характеризується своїми труднощами. Часто дані в транзакційних базах даних є сильно пов'язаними, і незважаючи на високе значення параметра достовірності правила і завдання булевого виразу, в результаті виходить дуже велика кількість асоціативних правил, за якими складно зробити якісь достовірні висновки і виявити цікаві закономірності. Тому надалі доцільно залучати підходи до оптимізації з обчислювальної точки зору першої підзадачі, а також шляхи скорочення кількості отриманих асоціативних правил, без можливої втрати даних.

**Результати аналізу зібраної інформації.** Виклад матеріалу цього пункту будемо здійснювати за наступною схемою: 1 – постановка конкретної задачі аналізу (мета), 2 – попередня обробка даних, 3 – структура вхідних даних для розв'язання поставленої задачі, 4 – інтерпретація отриманих результатів, висновки і рекомендації.

Таблиця 2

Крос-таблиця «Професія – Галузь народного господарства»

Професія	11	12	13	15	16	18	2	20	21	24	27	28	3	30	33	38	39	5	8
	зспо	зспо	зспо	зспо	зспо	зспо	зспс	зспс	зспо	зспс	зспс	зспо	зспс	зспо	зспо	зспо	зспо	зспо	зспс
Менеджер проектів в сфері ІТ	7		1	2	13	7	1	1	4		1			1	2	1	4	5	
Системний аналітик	7	1	1	2	21	1	2				1	1			5	1	2	3	2
Спеціаліст по ІР	9			1	9	10	8	1	1		4		2		1		1	1	1
Спеціаліст по ІС	7		1	4	13	2	6		1	3					1	2	2	8	1

**Задача 1.** В якій галузі народного господарства, науки чи виробництва найбільш затребувані професії?

В базі даних формується запит полів: «кодів респондента», «код галузі», «професія».

За допомогою платформи Deductor 5.2 будується крос-таблиця з агрегацією фактів, в якій відображується кількість респондентів певної галузі, які висловили своє судження про компетенції фахівців відповідної професії (табл. 2).

Отриману інформацію можна візуалізувати у вигляді діаграм, наприклад, наведених на рис. 2.

Отже, 1) фахівці з усіх чотирьох професій, що розглядаються, найбільш затребувані в галузі ІТ та торгівлі, причому, якщо у першій галузі потреба в системних аналітиках значно перевищує решту, то у другій – всі професії представлені досить поширено і майже рівномірно; з урахуванням останнього можна винести рекомендації щодо наявності у освітніх програмах підготовки фахівців

з зазначених професій вибіркових дисциплін, пов'язаних, наприклад, з обліком і аудитом або маркетингом; 2) спеціалісти з інформаційних ресурсів однаково затребувані у таких галузях, як ІТ, торгівля, реклама, невиробничі види побутового обслуговування; 3) спеціалісти з інформаційних систем, крім ІТ й торгівлі, задіяні також у галузях фінансів та кредиту, страхування, телекомунікацій; 4) професія менеджера проектів у сфері ІТ характерна і для галузей реклами, фінансів, медіа та інших (не представлених у базі даних).

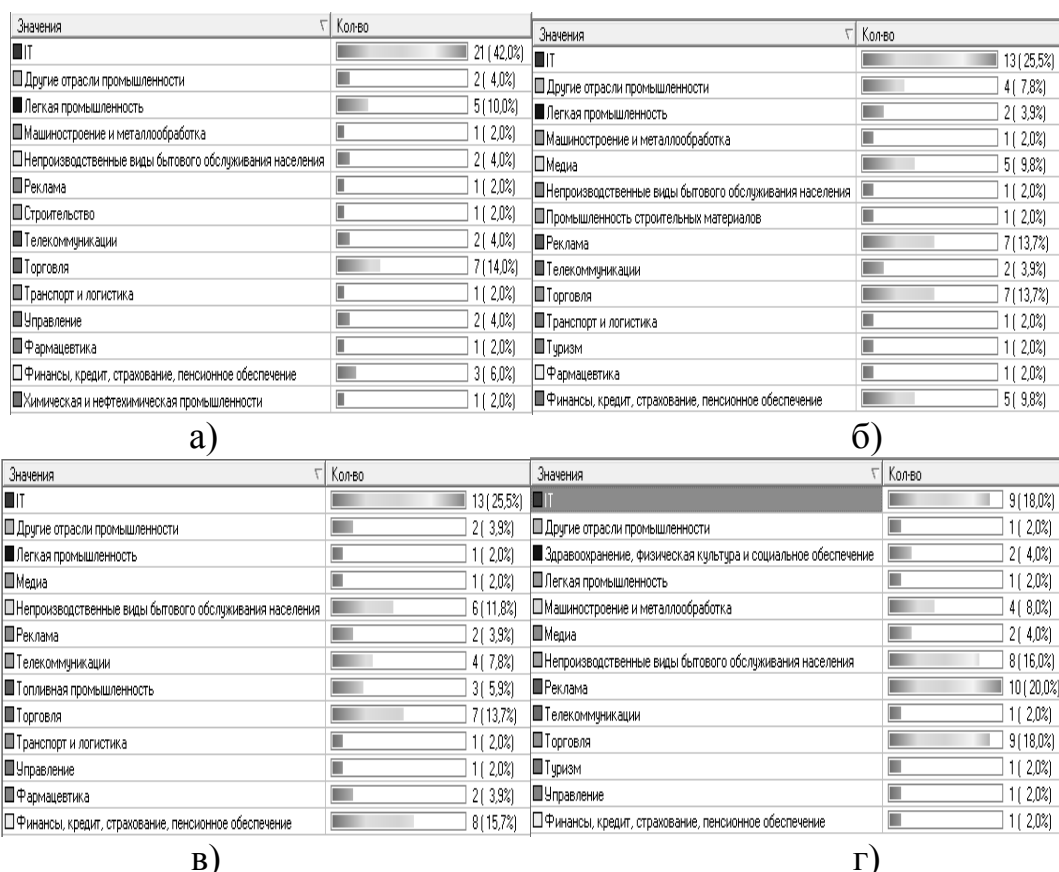


Рис. 2. Зацікавленість у фахівцях: а) – системний аналітик; б) – менеджер по проектах; в) – спеціаліст з ІР; г) – спеціаліст з ІС

**Задача 2.** Пошук наборів найбільш поширених компетенцій для кожної конкретної професії. Для перевірки коректності та адекватності математичної постановки задачі реалізуємо алгоритм її розв'язання також за допомогою аналітичної платформи для створення прикладних рішень Deductor 5.2. Наведемо деякі результати аналізу зібраної в БД інформації, не претендуючи на повноту викладу, а лише демонструючи можливість виявлення відомостей, корисних для складання навчальних програм за допомогою описаного підходу.

Вхідні дані – ідентифікатор респондента та шифр компетенції – були отримані шляхом запиту з бази даних лише тих записів, які містять вимоги до знань та вмінь фахівця з професії, що розглядається. В сценарії пошуку асоціативних правил було налаштовано такі параметри: мінімальна підтримка набору компетенцій 16%; максимальна підтримка набору компетенцій 87%; мінімальна кількість елементів в наборі 1; максимальна кількість елементів в наборі 5 (при

цьому максимальне число компетенцій, за які голосував респондент, дорівнює 8); мінімальна достовірність асоціативних правил 60%; максимальна підтримка асоціативних правил 90%.

За таких значень параметрів були виявлені: для професії Системний аналітик 25 популярних наборів компетенцій, для спеціаліста з ІС і Менеджера проєктів в галузі ІТ – по 25. Частина таких наборів компетенцій для системних аналітиків наведена на рис. 3. Нагадаємо, що популярні набори – це множини, що складаються з одного і більше елементів, які найбільш часто зустрічаються в транзакціях одночасно. Наскільки часто вони зустрічається у досліджуваному списку транзакцій, можна судити з їхньої підтримки.

Множеств: 25 из 39		Фильтр: Минимальная поддержка = 15.65		
№	№ Номер множества	аб. Элементы	Поддержка	
			Кол-во	%
1	37	Аналитические способности	8	16.00
		Внимательность к деталям		
		Ответственность		
2	36	Аналитические способности	8	16.00
		Внимательность к деталям		
		Знание Excel		
3	32	Знание Excel	10	20.00
		Умение работать с большими объемами информации		
4	27	Внимательность к деталям	8	16.00
		Уверенный пользователь MS Office, Microsoft Project		
5	26	Внимательность к деталям	9	18.00
		Ответственность		
6	25	Внимательность к деталям	9	18.00
		Знание Excel		
7	24	Аналитические способности	10	20.00
		Уровень английского - выше среднего		
8	23	Аналитические способности	8	16.00
		Умение работать с большими объемами информации		

Рис. 3. Найбільш популярні набори компетенцій за результатами побудови асоціативних правил для професії Системний аналітик

Використовуючи виявлені набори компетенцій, будуються дерева правил – дворівневі дерева за умовою або по слідству. При побудові дерева правил за умовою, на першому (верхньому) рівні знаходяться вузли з умовами, а на другому рівні – вузли зі слідством. Другий варіант дерева правил – дерево, побудоване по слідству. Тут на першому рівні розташовуються вузли зі слідством. Як приклад, на рис. 4 наведене дерево правил, побудоване на основі найпопулярніших наборів компетенцій для професії Спеціаліст з інформаційних систем. Тут – 5 правил. Їхня деталізація представлена на рис. 5, а інтерпретація може бути такою:

– якщо роботодавець вимагає від претендента на посаду спеціаліста з ІС аналітичних здібностей, то у 25,5% випадків разом з цим буде вимагати й уваги до деталей;

Правил: 5 из 5		Фильтр: Без фильтрации		Поддержка		Достоверность
№	Номер правила	Условие	Следствие	Кол-во	%	
1	1	Внимательность к деталям	Аналитические способности	6	11.76	85.71
2	2	Глубокие знания СУБД Oracle 1	Уровень английского - выше	6	11.76	60.00
3	3	Знание методологии и методик	Уровень английского - выше	7	13.73	87.50
4	4	Уверенный пользователь MS O	Уверенный пользователь ПК	9	17.65	64.29
5	5	Уверенный пользователь ПК	Уверенный пользователь MS	9	17.65	90.00

Рис. 4. Правила щодо складання наборів компетенцій спеціаліста з інформаційних систем

– якщо роботодавець вважає, що рівень володіння англійською мовою у претендента на посаду спеціаліста з ІС має бути вище за середній, то у 12% разом з цим претендент повинен мати глибокі знання СУБД Oracle, мов програмування SQL, PL/SQL і т.д.

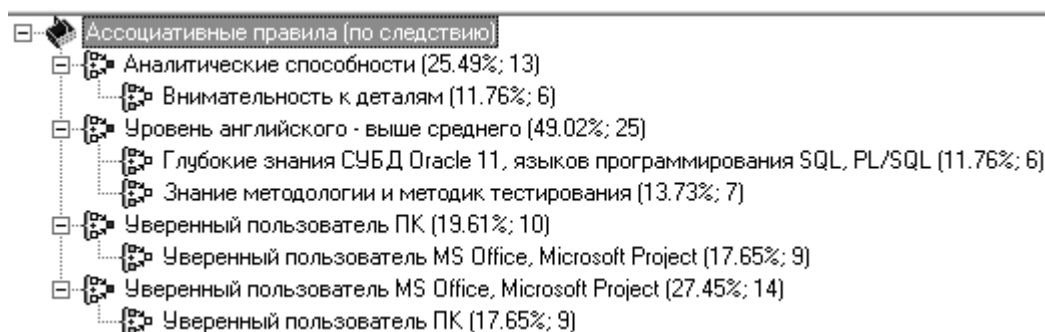


Рис. 5. Деталізація дерева правил щодо складання наборів компетенцій спеціаліста з інформаційних систем

На рис. 6 представлена деталізація деяких правил щодо складу компетенцій системного аналітика за вимогами роботодавця.

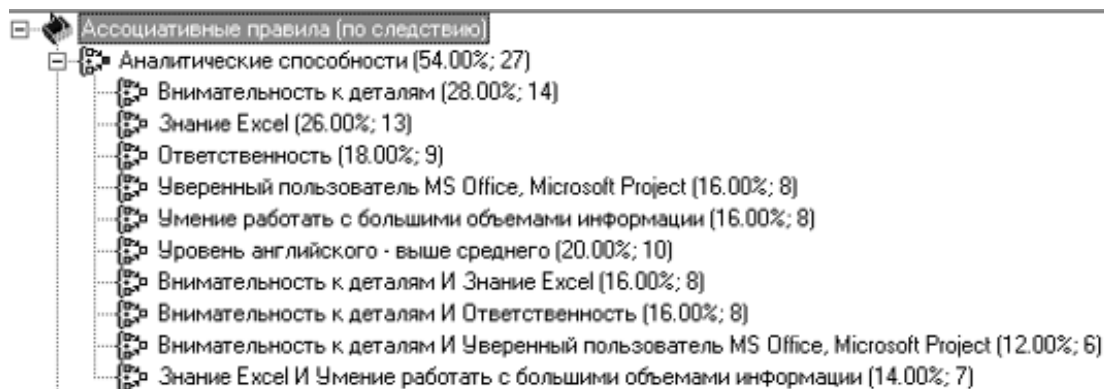


Рис. 6. Деталізація правил складання наборів компетенцій

Загалом, за вказаних вище параметрах, побудовано правил типу «умова – наслідок»: для системного аналітика – 20, для Менеджера по проектах в галузі ІТ – 6. За допомогою дерева правил, аналізуючи склад найпопулярніших набо-

рів компетенцій для всіх професій, що розглядаються, можна також простежити взаємозв'язок між цими професіями.

В табл. 3 наведені спільні компетенції для кандидатів на посади системного аналітика, спеціаліста з інформаційних систем, спеціаліста з інформаційних ресурсів і менеджера по проектах в галузі ІТ, які вдалося виявити за допомогою аналізу зібраної інформації щодо вимог роботодавців. Такі зведені дані демонструють спорідненість вищезгаданих професій і можуть бути корисними під час розробки або порівняння освітніх програм підготовки фахівців.

За допомогою методів Data Mining можна також виявити і ті компетенції, які на думку роботодавців властиві лише одній професії. І отже, здійснивши відсів тих знань і умінь, які мають підтримку, меншу за деякий поріг, можна сформулювати перелік компетенцій, що визначають специфіку окремої професії.

**Висновки.** Представлені два підходи до систематизації і аналізу результатів опитування зацікавлених осіб про компетенції, властиві для тих чи інших професій. Перший підхід полягає у відборі найбільш затребуваних компетенцій за сумарною виваженою оцінкою кожної з них. Другий підхід включає пошук наборів компетенцій, що часто зустрічаються, і складання бази асоціативних правил, яка дає більш повну картину про компетенції, властиві для певної професії, тому що дозволяє виявити непряму інформацію про взаємозв'язки між самими компетенціями (на думку осіб, які беруть участь в опитуванні).

Практика показує, що застосування методів Data Mining при аналізі результатів голосування про знання і уміння, якими повинен володіти той чи інший фахівець, має низку переваг. По-перше, можна враховувати не тільки ті варіанти голосувань, при яких досить велика кількість респондентів голосують за одну-дві певні компетенції, а й цілі блоки компетенцій, притаманних даним професії з точки зору осіб, зацікавлених в результатах голосування. По-друге, якщо у великій базі даних, де зібрана інформація за найрізноманітнішими професіями про компетенції, необхідні з точки зору зацікавлених осіб, реальна ситуація, що один і той самий достатньо великий набір знань і умінь є властивим, з точки зору респондентів, для декількох професіях, то можна говорити про те, що ці професії взаємозамінні, або споріднені. По-третє, зіставляючи галузі науки і виробництва, до яких належать респонденти і галузі професій, судження про компетенції яких винесли відповідні зацікавлені особи, можна виявити професії, асоційовані з різними галузями. Це в свою чергу дозволяє вносити рекомендації по наповненню освітніх програм підготовки фахівців.



Таблиця 3

Спільні компетенції для професій Системний аналітик, Спеціаліст з ІС, Спеціаліст з ІР, Менеджер по проектах в галузі ІТ

Тип компетенції	Найменування компетенції	Професії
Професійно-виробнича (ПВ)	Рівень англійської - вище середнього Уміння складати і працювати з тестовою документацією Уміння працювати з ІС, Впевнений користувач ПК Впевнений користувач MS Office, Microsoft Project Предметна область, Розуміння принципів SEO Розуміння основних принципів розробки ПЗ Основи HTML / PHP / MySQL / JavaScript Основи менеджменту, Знання CRM системи, Знання Excel Грамотна усна і письмова мова Глибокі знання з e-commerce системами Базові знання SQL, Web програмування	Спеціаліст з інформаційних систем (C_IP)  Спеціаліст з інформаційних ресурсів (C_IC)  Системний аналітик (CA)
Особистісна	Цілеспрямованість, багатозадачність, стресостійкість, порядність Уміння ставити завдання і досягати поставлених цілей Уміння працювати з великими обсягами інформації Працьовитість, самодисципліна, Відповідальність, ініціативність, акуратність, Спрямованість на результат, Креативний підхід знаходити власні шляхи розв'язання Уважність до деталей, Здатність швидко навчатися Аналітичні здібності, Активна життєва позиція	Менеджер проектів в галузі ІТ (МП)
Соціальна	Високий рівень комунікативних навичок	
Управлінська	Уміння організувати роботу команди і працювати в ній	
Професійно-виробнича	Розуміння структури веб розробки Уміння працювати з проектною документацією (BRD / FSD / UMD / MRD) Уміння працювати з Redmine / JIRA Основи QA (тестування), Навички управління проектами Навички аналітичної роботи (аналіз клієнтської бази, маркетингові дослідження, профіль клієнта, сегментація бази, визначення ЦА) Володіння системою моделювання бізнес процесів BPMN	Системний аналітик (CA)  Менеджер проектів в галузі ІТ (МП)
Особистісна	Впевненість в собі, Логічне мислення	
Соціальна	Уміння проводити презентації, Вміння вести переговори Здатність створювати і підтримувати довгострокові відносини з клієнтом	
Управлінська	Уміння контролювати і доводити робочі процеси до їх повного завершення	C_IC CA МП
ПВ	Навички роботи з ERP системами	
Професійно-виробнича	Навички в тестуванні API, web і mobile додатків Знання Unix shell / awk Глибокі знання СУБД Oracle 11, мов програмування SQL, PL/SQL	C_IC CA
ПВ	Знання ринку реклами Знання графічних редакторів	C_IP МП
Особистісна	Енергійність, Ерудованість	
Особистісна	Уміння працювати з пошуковими системами	C_IP, CA
ПВ	Web дизайн	C_IP, C_IC

Надалі буде корисним і важливим доповнити базу даних результатами анкетування інших респондентів, які є зацікавленими особами або можуть виступати експертами. Для того, щоб не змушувати голосуючого переробляти при опитуванні велику кількість інформації, при складанні варіантів можливих компетенцій потрібно їх структурувати (агрегувати, вибудувати ієрархію), з

можливістю розшифровки тієї чи іншої агрегованої компетенції. Крім того, ефективність роботи з аналізу динаміки зміни якості освіти багато в чому залежить від того, наскільки правильно структурована сукупність компетенцій, наскільки вона піддається операціоналізації, представляється у вигляді деякої системи показників, що піддаються вимірюванню або експертному оцінюванню.

#### Перелік посилань

4. Communication from the Commission to the European Parliament, The Council, The European Economic And Social Committee And The Committee of The Regions. Supporting growth and jobs – an agenda for the modernisation of Europe's higher education systems. – <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52011DC0567>
5. Політика забезпечення якості вищої освіти. – Дніпропетровськ, НГУ. – 2016. – [http://www.nmu.org.ua/ua/content/infrastructure/structural\\_divisions/science\\_met\\_cent/skladov-i-sistemi-zabezpechennya-yakosti-osviti-ngu](http://www.nmu.org.ua/ua/content/infrastructure/structural_divisions/science_met_cent/skladov-i-sistemi-zabezpechennya-yakosti-osviti-ngu)
6. Развитие системы обеспечения качества высшего образования в Украине: информационно-аналитический обзор / Укладачі: Добко Г., Золотарьова І., Калашнікова С., Ковтунець В., Курбатов С., Линьова І., Луговий В., Прохор І., Рашкевич Ю., Сікорська І., Таланова Ж., Фініков Т., Шаров С.; за заг. ред. С. Калашнікової та В. Лугового. – Київ : ДП «НВЦ «Пріоритети», 2015. – 84 с.
7. Построение программной логики при моделировании связанных компетентностных систем (Украины, России, Европейского Союза)/ Всеволод Чернышенко// Universität Koblenz-Landau. 15 июля 2015
8. Aleksieiev M.A. Applying data mining methods for the analysis and improving of composition of professional educational standards / M.A.Aleksieiev, V.S. Chernyshenko, L.S. Koriashkina // Abstracts of the X International Conference «Modern information and communication technologies on a transport, in industry and education» (Dnipro, 14.12.2016 – 15.12.2016). – ДІПТ, 2016. – P. 135 – 136.

#### ABSTRACT

**Purpose.** Development of methods and algorithms for analyzing the quality of higher education, researching the labor market of graduates, creating mechanisms for comparing students' knowledge and skills they receive at the university with competencies that specialists in relevant professions should have in terms of employers

**The methodology** of research consists on collection and systematization of relevant information, its study and analysis using methods of mathematical modeling and Data Mining.

**Findings.** The MS Access database "Employers' requirements" has been developed based on information from hundreds of employers' letters containing the lists of knowledge and skills that applicants should have for one of the following positions: system analyst, project manager, information resource specialist, expert in information systems. An analysis of the requirements of employers is conducted and recommendations on the content of the relevant educational programs are made.

**The originality.** Two approaches to organizing and analyzing the results of a survey of stakeholders' competence peculiar to certain professions are presented. The first approach is the selection of the most sought after competencies on a total weighted assessment of each. The second approach involves searching sets of competences, common and associative drafting rules base that gives a more complete picture of the competencies inherent to a profession because it allows to detect indirect information about the relationship between the most competences (according to those who are participated in the survey).

**Practical implications.** When analyzing the results of voting on the knowledge and skills that should have a particular specialist the methods of Data Mining allowed: to take into account not only the options of voting, in which a fairly large number of respondents vote for one or two specific competence, but whole blocks of competencies inherent to this profession; to determine interchangeable or related professions; to reveal the connection between the professions and the branches of science and production, where these professions are most in demand; to make recommendations for filling the educational training programs.

**Keywords:** *quality training, system analysis, employer's requirements analysis, Data Mining*

УДК 681.5:621.314.57

© С.М. Ткаченко

### **ДИНАМІЧНЕ НАЛАГОДЖЕННЯ ВИХОДІВ ТА ВХОДІВ ТИПУ «ТАК-НІ» У ТЕГОВІЙ ПАМ'ЯТІ ПРОМИСЛОВОГО КОНТРОЛЕРА**

© S. Tkachenko

### **DYNAMIC DEBUGGING OF THE OUTPUTS AND INPUTS OF THE "YES-NO" TYPE IN THE TAG MEMORY OF COMPTROLLERS**

Розроблено метод непрямой побітової адресації тегових областей пам'яті виходів та входів «Так-Ні» промислового контролера для динамічного налагодження і перепризначення каналів вводу і виводу сигналів технологічного устаткування без програмування.

Разработан метод непрямой побитовой адресации теговых областей памяти выходов и входив «Да-нет» промышленного контроллера для динамической отладки и переназначения каналов ввода и вывода сигналов технологического оборудования без программирования.

**Вступ.** Для багатьох підприємств України можливості використання промислових контролерів обмежені через високу вартість впровадження та експлуатації програмно-технічних засобів. За умови їх використання у порядку, передбаченому фірмою-розробником, кожний сигнал має бути заведений на окремий вхід чи вихід, а кожна інформаційна ознака, що передається або отримується від засобів відображення SCADA, повинна мати свій ліцензований тег, тобто, за перекладом з англійської, «ознаку». Вартість такої ліцензії, в залежності від потреби у тегах, знаходиться, приблизно, у межах від 300 до 4000 євро для різних виробників обладнання[1]. Ряд крупних виробників, таких, наприклад, як Siemens або Ge-Fanuc, для обміну дозволяють використовувати додаткові засоби, такі як масиви, строки та гранд-теги. Вони дозволяють в один ліцензований тег включити досить великі обсяги даних, порядку сотень або й тисяч байт, але тут універсального підходу не існує і робота зі складання та розкладання такої структури даних лягає на програміста.