

УДК 004.65:004.82

Дворецький М.Л.

Інтелектуальний аналіз даних в 1С:8.0

В статті розглянуто реалізацію задач пошуку асоціацій, пошуку послідовностей, задачу кластеризації та побудову дерева рішень засобами вбудованої мови системи 1С:8.0. Розглянуто основні об'єкти системи аналізу даних і прогнозування, та запропоновано підхід щодо створення бази знань, що поєднує у собі дані із результатами аналізу для подальшої побудови моделей прогнозу.

In article is discussed a realization of data mining tasks, such as associative roles search, sequence analysis, decision trees and cluster analysis by means of 1C:8.0 programming language. Also are considered main data analysis system objects and prediction objects. Is given the way of knowledge base creation with join results of data analysis for the following prediction model.

Загальна постановка проблеми та її зв'язок з науково-практичними задачами. Механізм аналізу даних і прогнозування дозволяє реалізовувати в прикладних рішеннях різні засоби для виявлення закономірностей, що звичайно ховаються за великими обсягами інформації [1].

Загальну схему роботи механізму аналізу і прогнозування даних наведено на рис. 1.

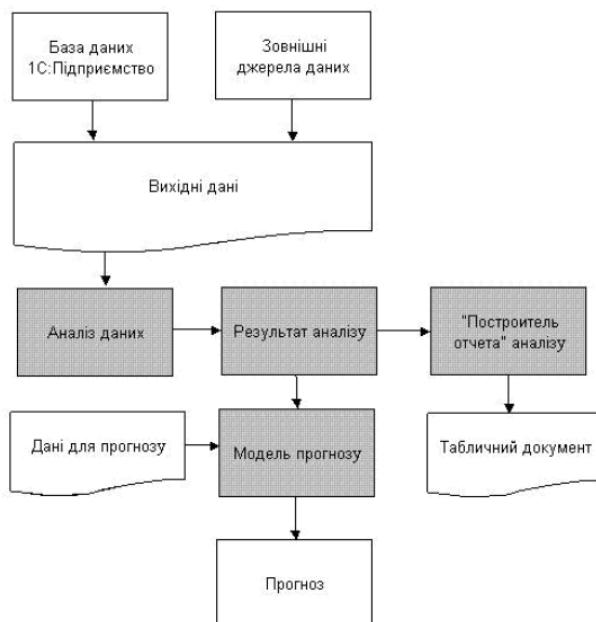


Рис.1. Загальна схема роботи механізму аналізу і прогнозування даних в 1С 8.0

Застосовуючи до вихідних даних один з видів аналізу, можна одержати результат аналізу. Результат аналізу являє собою деяку модель поведінки даних. Результат аналізу може бути відображенний у підсумковому документі чи збережений для

подальшого використання [2].

У механізмі аналізу даних, що розглядається, реалізовано такі типи аналізу даних, як:

- пошук асоціацій;
- пошук послідовностей;
- кластерний аналіз;
- дерево рішення.

Схематично процес проведення аналізу по вищезазначенним типам наведено на рис.2-5 [3].

Тип аналізу **пошук асоціацій** здійснює пошук комбінацій об'єктів, чи значень характеристик, що зустрічаються часто разом. З погляду даного типу аналізу стовпчики джерела можна поділити на:

- невикористовувані. Ігноруються аналізом;
- об'єкт. Дані з цього стовпчика використовуються як об'єкти (чи події) проведеного аналізу. Виходячи зі значенням даного стовпчика значення іншого стовпчика (Елемент) відносяться до однієї асоційованої групи;
- елемент. Дані з цього стовпчика використовуються для одержання стійких груп значень, побудови асоціативних правил.

Крім настроювання типів колонок, на результат проведеного аналізу впливають такі параметри аналізу:

- Мінімальний Відсоток Випадків. Визначається мінімальний відсоток випадків, у яких повинна зустрічатися комбінація елементів. Групи, у яких дане значення менше зазначеного, не попадають у результат аналізу;
- Мінімальна Достовірність. Показує мінімальне значення відсотка випадків, коли правило витримується;
- Мінімальна Значимість. Групи, у яких дане значення менше зазначеного, не попадають у результат аналізу;
 - Тип Відтинання Правил. Один з варіантів системного перерахування;
 - Тип Відтинання Правил Асоціацій: Надлишкові (відтинати надлишкові правила), Покриті (відтинати правила, покриті іншими правилами).

У результаті виконання аналізу одержуємо:

- інформацію про дані (кількість об'єктів, кількість елементів, середня кількість елементів в об'єкті, кількість знайдених груп, кількість знайдених правил асоціацій);
- знайдені групи елементів. Указується склад групи, кількість випадків, відсоток випадків, у якому ця група зустрічається;
- знайдені асоціативні правила. Указується вихідний склад елементів, наслідок (склад елементів), відсоток випадків, вірогідність, значимість правила.

Пошук послідовностей дозволяє виявити ланцюжки виникаючих подій (шаблони послідовностей). Він може використовуватися тоді, коли одним з важливих аналізованих показників є послідовність настання подій у часі. [3].

З погляду даного типу аналізу стовпчики джерела можна поділити на:

- Невикористовувані. Ігноруються аналізом;
- Послідовність. Дані з цього стовпчика використовуються для аналізу як об'єкт події послідовності. За значенням даного стовпчика аналіз і асоціює дані з одним ланцюжком подій;

- Елемент. Дані з цього стовпчика використовуються як елементи послідовності;
- Час. Саме по даному стовпчику визначається час настання події. Наявність даного стовпчика обов'язкова при проведенні даного типу аналізу;

Крім настроювання типів колонок, на результат проведеного аналізу впливають такі параметри аналізу:

- Мінімальний Відсоток Випадків. Мінімальний відсоток послідовностей, у яких спостерігається знайдений шаблон послідовності;
- Мінімальний Інтервал. Ознака установки мінімального інтервалу послідовності (повинна бути встановлена одиниця виміру інтервалу, кратність);
- Максимальний Інтервал. Ознака установки максимального інтервалу послідовності (повинна бути встановлена одиниця виміру інтервалу, кратність);
- Інтервал Еквівалентності Часу. Ознака установки інтервалу еквівалентності часу (повинна бути встановлена одиниця інтервалу еквівалентності часу, її кратність);
- Мінімальна Довжина – Мінімальна довжина шуканих послідовностей;
- Пошук За Ієрархією. Ознака здійснення пошуку за ієрархією (поширюється на стовпчики з типом Елемент).

Основним результатом аналізу є знайдені шаблони послідовностей. Цей шаблон містить таку інформацію:

- склад шаблону послідовності;
- кількість випадків, у якому спостерігалася дана послідовність;
- максимальні інтервали між подіями (у випадку, якщо подій 2, то інтервал один);
- мінімальні інтервали між подіями (у випадку, якщо подій 2, то інтервал один);
- відсоток випадків, коли дана послідовність виконалася;
- середні інтервали між подіями (у випадку, якщо подій 2, то інтервал один).

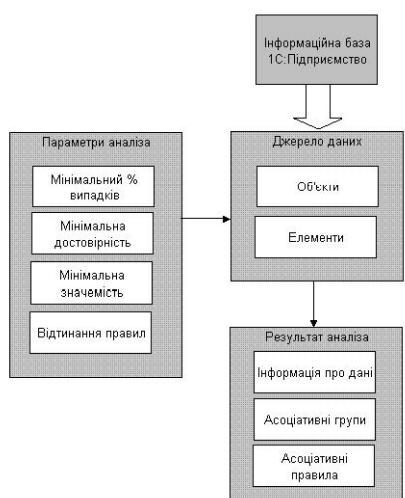


Рис. 2. Тип аналізу пошук асоціацій

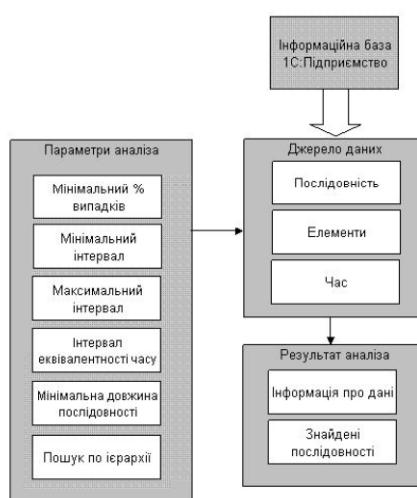


Рис. 3. Тип аналізу пошук послідовностей

За допомогою типу аналізу **дерево рішення** можна одержати причинно-наслідкову ієрархію умов, що приводить до визначеного рішення.

З погляду даного типу аналізу джерела можна поділити на:

- Невикористовувану;

- Вхідну;
- Прогнозовану.

Використовувані параметри аналізу:

- Мінімальна Кількість Випадків – мінімальна кількість елементів у вузлі;
- Максимальна Глибина – максимальна глибина дерева;
- Тип Спрощення – тип спрощення дерева рішень.

У результаті проведеного аналізу можна одержати:

- дерево рішень;
- помилки класифікації.

Кластерний аналіз – математична процедура багатомірного аналізу, що дозволяє на основі безлічі показників, що характеризують ряд об'єктів, згрупувати їх у групи (кластери) таким чином, щоб об'єкти, що входять в один кластер, були більш однорідними, подібними в порівнянні з об'єктами, що входять в інші кластери.

В основі даного аналізу лежить обчислення відстані між об'єктами. Саме виходячи з відстаней між об'єктами і виробляється їхнє групування по кластерах. Визначення відстані може проводитися різними способами (по різних метриках) [1]. Підтримуються такі метрики:

- Евклідова метрика;
- Евклідова метрика в квадраті;
- Метрика міста;
- Метрика домінування.

Після визначення відстаней між об'єктами може використовуватися один з декількох алгоритмів розподілу об'єктів по кластерах. Підтримуються наступні методи кластеризації:

- Близкий зв'язок;
- Далекий зв'язок;
- Центр ваги;
- "k середніх".

Стовпчики джерела визначаються як вхідні або невикористовуванні. Слід зазначити, що всі значення стовпчиків містяться в системному перерахуванні ТипКолонкиАналізуДанихКластеризація. У цьому перерахуванні значень більше (ніж невикористовувані і вхідні), але інші значення використовується при побудові прогнозів.

Аналіз проводиться відповідно до встановлених параметрів аналізу.

Подальше використання результату аналізу полягає в тому, що на його основі може бути створена модель прогнозу, що дозволяє прогнозувати поведінку нових даних відповідно до наявної моделі [4].

Схематично взаємозв'язок основних об'єктів механізму аналізу даних і прогнозування наведено на рис 6.

АналізДаних – об'єкт, що безпосередньо виконує аналіз даних. Йому встановлюється джерело даних, задаються параметри і вихідні дані. Результатом роботи даного об'єкта є результат аналізу даних, причому для кожного типу аналізу існує свій об'єкт для роботи з результатом аналізу (РезультатАналізуДанихЗагальнаСтатистика; РезультатАналізуДанихПошукАсоціацій; РезультатАналізуДанихПошукПослідовностей; РезультатАналізаДанихДревоРішень; РезультатАналізуДанихКластеризація)

Результат аналізу даних — спеціальний об'єкт, що містить інформацію про

результат аналізу. Для кожного виду аналізу передбачений свій результат. Наприклад, результатом аналізу даних дерева рішень буде об'єкт типу РезультатАналізуДанихДеревоРішень. Надалі результат може бути виведений у табличний документ, може бути використаний для створення *моделі прогнозу*.

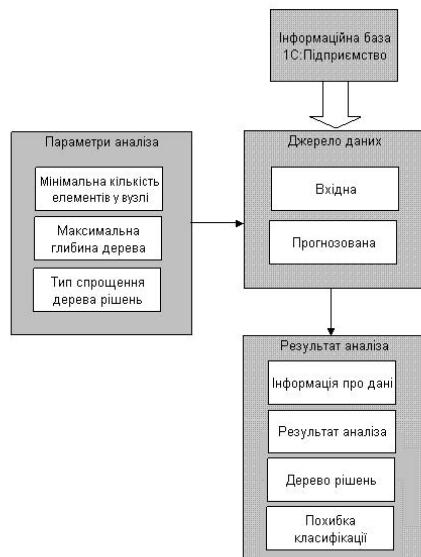


Рис. 4. Тип аналізу дерево рішень

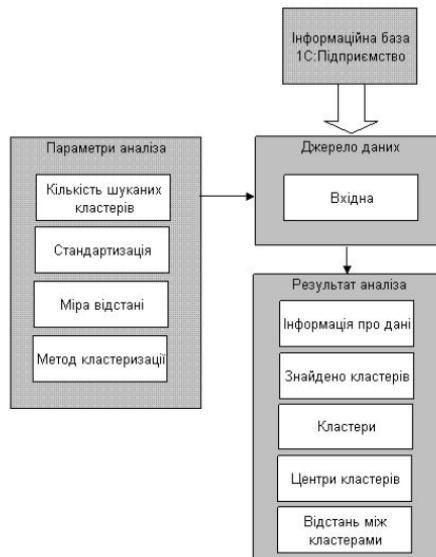


Рис. 5. Тип аналізу кластерний аналіз

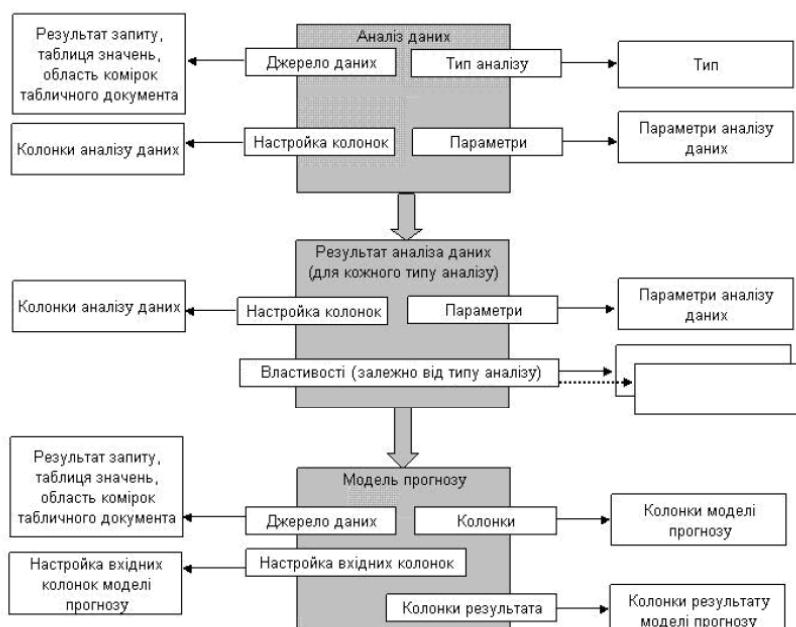


Рис. 6. Взаємозв'язок основних об'єктів механізму аналізу даних і прогнозування

Модель прогнозу — спеціальний об'єкт, що дозволяє виконувати прогноз на підставі вхідних даних. Тип моделі прогнозу залежить від типу результата аналізу даних. Наприклад, модель, створена для типу аналізу пошук асоціацій, буде мати тип МодельПрогнозуПошукАсоціацій. На вхід моделі прогнозу передається джерело даних для прогнозу. Результатом є таблиця значень, що містить прогнозовані значення [3].

Серед недоліків підсистеми слід зазначити таке. Модель прогнозу може використовувати лише один результат аналізу даних, тобто неможливе накопичення

різних результатів аналізу із подальшим порівнянням та обранням більш оптимального. Крім того, неможливе прогнозування одразу за декількома типами аналізу даних. Тому постає актуальним створення єдиної бази знань, у якій би накопичувалися результати аналізу даних у вигляді набору правил із впровадженням алгоритмів відтинання покритих та збиткових правил.

Огляд публікацій та аналіз невирішених питань. Концепція знань, що поєднують у собі риси процедурної та декларативної інформації, виникла досить давно. Існує цілий ряд публікацій, присвячених моделям представлення знань в інтелектуальних системах [1,3,7-8], та методам роботи зі знаннями [9-11]. Також були проаналізовані механізми аналізу прогнозування, реалізовані у програмному продукті 1С:8.0 на базі алгоритмів Data mining [2-6].

Однак задача побудови бази знань у вигляді набору правил шляхом застосування до навчаючої вибірки алгоритмів Data mining залишається невирішеною, а отже актуальною.

Метою досліджень є підвищення якості прийняття управлінських рішень шляхом розробки на базі конфігурації 1С:8.0 “Аналіз даних” модуля побудови бази знань із застосуванням механізмів аналізу даних та модуля прогнозу.

Результати досліджень. Якщо розглянути задачу прогнозування загалом, то в ній можна виділити дві важливі частини:

- навчання моделі на який-небудь вибірці даних;
- використання навченої моделі для роботи з фактичними даними для одержання якого-небудь прогнозу.

Коли мова йде про який-небудь експертний висновок, теж можна говорити про ці етапи. Перш ніж давати висновок, прогноз, людина-експерт тривалий час учиться приймати правильні рішення, вивчає закономірності в цікавлячій його прикладній області, що його цікавить і т.п.

Далі розглянемо модель прогнозу в конфігурації 1С:8.0 “Аналіз даних”. У даному випадку у ролі експерта виступає модель прогнозу. Даний об'єкт може бути отриманий тільки на підставі проведеного аналізу даних. Можна сказати, що при проведенні аналізу виробляється навчання майбутньої моделі прогнозу (вірніше сказати, вона створюється після проведення аналізу на деякій аналітичній вибірці). Після створення в модель прогнозу може бути передана деяка фактична вибірка даних, за якою (на підставі закономірностей, отриманих під час аналізу) буде побудований прогноз.

Схематично модель прогнозу наведено на рис. 7.

Створення моделі прогнозу доступно для всіх типів аналізу. Модель прогнозу має визначені властивості, за допомогою яких можна керувати проведеним прогнозом. При роботі з колекцією стовпчиків можна вказувати такі типи стовпчиків (значення системного перерахування ТипКолонкиМоделіПрогнозу):

- Вхідна;
- КолонкаДжерелаДаних;
- Прогнозована.

Слід зазначити, що для кластерного аналізу в колекції Стовпчики визначена властивість ДодатковіПараметри (визначається вага стовпчика і значення нормалізації).

Колекція Настройкавхіднихколонок містить опис відповідності колонок джерела даних (за якою будується прогноз) стовпчикам моделі прогнозу. Колекція Колонкирезультата містить стовпчики, виведені в результат прогнозу.

Найбільш популярними засобами представлення знань є продукції. Продукції, з

одного боку, близькі до логічних моделей, що дозволяє організовувати на них ефективні процедури висновку, а з іншого боку, більш наочно, ніж класичні логічні моделі, відображають знання. У них відсутні тверді обмеження, характерні для логічних моделей, що дає можливість змінювати інтерпретацію елементів продукції.

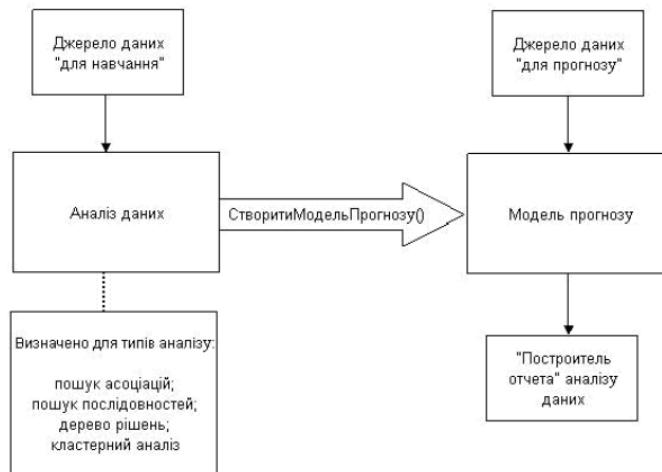


Рис. 7. Модель прогнозу

У загальному вигляді під продукцією розуміється вираження такого виду: (i); Q;P;A⇒B;N.

Тут i – ім'я продукції, за допомогою якого дана продукція виділяється з усієї множини продукції. Як ім'я може виступати деяка лексема, що відбиває суть даної продукції, чи порядковий номер продукції у її множині, що зберігається в пам'яті системи.

Елемент Q характеризує сферу застосування продукції. Поділ знань на окремі сфери дозволяє заощаджувати час на пошук потрібних знань. Такий же поділ на сфери в базі знань доцільний і при використанні для представлення знань продукційних моделей.

Основним елементом продукції є її ядро: A⇒B. Інтерпретація ядра продукції може бути різною і залежить від того, що стоїть ліворуч і праворуч від знака секвенції ⇒. Звичайно розшифровка ядра продукції виглядає так: ЯКЩО A, ТО B, більш складні конструкції ядра допускають у правій частині альтернативний вибір, наприклад, ЯКЩО A, ТО B₁, ІНАКШЕ B₂. Секвенція може сприйматися в звичайному логічному змісті як знак логічного слідування B із істинного A (якщо A не є істинним виразом, то про B нічого сказати не можна). Можливі й інші інтерпретації ядра продукції, наприклад, A описує деяку умову, необхідну для того, щоб стало можливим B.

Елемент P є умова застосовності ядра продукції. Звичайно P являє собою логічний вираз (як правило, предикат). Коли P приймає значення "істина", ядро продукції активізується. Якщо P "хибне", то ядро продукції не може бути використане.

Елемент N описує постумову продукції. Вони актуалізуються тільки в тому випадку, якщо ядро продукції спрвджується. Постумова продукції описує дії і процедури, які необхідно виконати після реалізації B. Виконання N може відбуватися відразу після реалізації ядра продукції.

Якщо в пам'яті системи зберігається деякий набір продукцій, то вони утворять систему продукцій. У системі продукцій повинні бути задані спеціальні процедури

керування продукціями, за допомогою яких відбувається актуалізація продукцій і вибір для виконання тієї чи іншої продукції з числа актуалізованих.

На рис. 8. наведено схему взаємодії підсистеми аналізу даних із базою знань через розроблений модуль представлення знань. Дані отримуються від об'єктів системи РезультатАналізуДанихПошукАсоціацій, РезультатАналізуДанихПошукПослідовностей, РезультатАналізуДанихДеревоРішень та РезультатАналізуДанихКластеризація та представляються у вигляді набору продукцій у базі знань. Модуль прогнозу дозволяє будувати модуль поведінки системи шляхом звернення до бази знань без очікувань результатів аналізу, що був попередньо проведений. Запропонована схема представлення результатів аналізу та побудови прогнозу дозволяє поєднувати у одній базі знань продукції, отримані з різних джерел даних та підвищує швидкість отримання прогнозу.

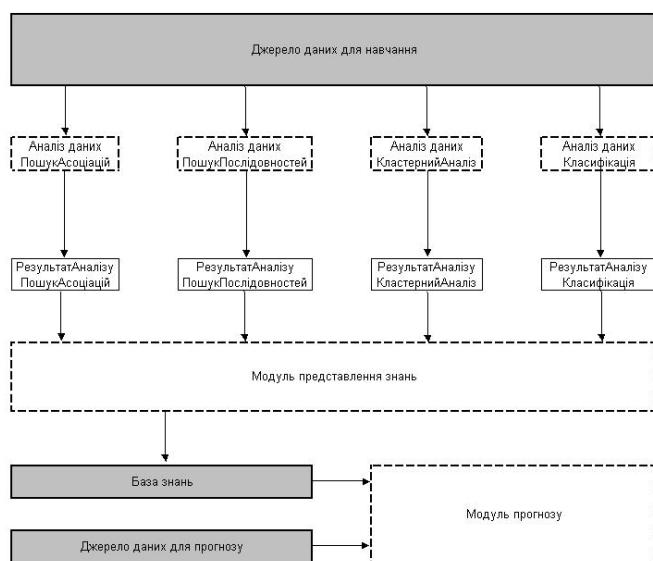


Рис. 8. Схема взаємодії підсистеми аналізу даних із базою знань та модулем прогнозу

Розроблені модулі представлення знань та прогнозу знаходяться на етапі тестової експлуатації на підприємстві оптової торгівлі ТОВ „Бестгроссер”, м.Миколаїв, у складі конфігурації „1С:Управління підприємством 8.0 для України”.

Висновки та перспективи подальших досліджень. Представлення даних, отриманих від об'єктів системи РезультатАналізуДанихПошукАсоціацій, РезультатАналізуДанихПошукПослідовностей, РезультатАналізуДанихДеревоРішень та РезультатАналізуДанихКластеризація в результаті проведення одного із типів аналізу у вигляді набору продукцій у базі знань дозволяє будувати модуль поведінки системи шляхом звернення безпосередньо до бази знань та джерела даних прогнозу. Запропонована модель представлення результатів аналізу та побудови прогнозу дозволяє поєднувати в одній базі знань продукції, отримані з різних джерел даних та підвищує швидкість отримання прогнозу.

Література

1. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – Питербург, 2004. – 336 с.
2. Габец А.П., Гончаров Д.И., Козырев Д.В., Кухлевский Д.С., Радченко М.Г. Профессиональная разработка в системе

- 1С:Предприятие 8. – Москва: 1С-Паблишинг, 2006. – 808 с.
3. 1С:Предприятие 8.0. Управление производственным предприятием для Украины. Общая концепция системы. - Москва: 1С-Паблишинг, 2005. – 298 с.
4. Щумелов Е.Л. Информационные технологии оптовой торговли: Практическое пособие. – Москва: 1С-Паблишинг, 2003. – 168 с.
5. Митичкин С.А. Разработка в системе 1С:Предприятие 8.0. – Москва: 1С-Паблишинг, 2003. – 413 с.
6. Радченко М.Г. 1С:Предприятие 8.0: Практическое пособие разработчика. – Москва: 1С-Паблишинг, 2004. – 656 с.
7. Гайдышев И. Анализ и обработка данных: Специальный справочник – СПб.: Питер, 2001. – 752 с.
8. Гарсия-Молина Г., Ульман Дж.Д., Уидом Дж. Системы баз данных. Полный курс. – М.: Изд. дом "Вильямс", 2003. – 1088 с.
9. Сошников Д.В., Дубовик С.Е. Синтез баз знаний для управления интеллектуальными бизнес-процессами на основе структурных функциональных моделей. <http://www.soshnikov.com/publications/IntBusinessProcess.pdf>
10. Морозов М.Н. Курс лекций по дисциплине "Системы искусственного интеллекта" Представление знаний в интеллектуальных системах (часть 2). – <http://khpi-iip.mipk.kharkiv.edu/library/ai/conspai/05.html>
11. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – Спб.: – Питер, 2001. – 384 с.