УДК 519.2:621.391

В.В. Краснопрошин, Г. Шаках, А.Н. Вальвачев

ТЕХНОЛОГИЯ ПОСТРОЕНИЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ РАСПРЕДЕЛЕННЫХ РАЗНОРОДНЫХ ЗНАНИЙ

Рассматриваются вопросы построения систем поддержки принятия решений на основе гетерогенных распределенных источников. Предложена модель системы и стратегия ее реализации для получения результата, релевантного исходной задаче. Особое внимание уделяется характеристике программных компонентов, которые признаны в настоящее время лидерами индустрии обработки знаний.

Введение

Успех деятельности субъектов экономики существенным образом зависит от правильности решений, принимаемых на различных уровнях управления. Эффективность решений определяется точностью, полнотой и актуальностью используемых знаний, в качестве которых рассматривается любой доступный информационный ресурс (например, базы знаний, базы данных, опыт экспертов). В настоящее время знания в задачах принятия решений (ЗПР) приобретают статус интеллектуального капитала, определяющего основные конкурентные преимущества субъекта рыночной экономики. Соответственно возрастает роль процессов добычи, хранения, своевременного обновления и использования знаний. Комплексное решение этих проблем осуществляется в рамках систем поддержки принятия решений (СППР) [1–3].

Эффективность СППР зависит от качества используемых программно-информационных ресурсов и соответствия механизма решения реалиям окружающей среды. Существующие СППР созданы в относительно стабильных экономических условиях для организационных структур с вертикальной системой управления и статическими бизнес-процессами. Для принятия решений используются хорошо формализованные предметные области (ПО), сформированные на основе знаний, полученных из ограниченного числа локальных источников, в качестве которых выступают эксперты и базы данных (БД), жестко встроенные (физически или логически) в архитектуру СППР [1, 2].

Процессы глобализации привели к политической, экономической и социальной нестабильности в мире. Пытаясь приспособиться к динамическому окружению, субъекты экономики приобрели ряд новых свойств, наиболее существенными из которых являются: горизонтальная система управления, динамические бизнес-процессы, территориальное распределение персонала [3]. Соответственно изменилась природа и масштаб ЗПР, они приобрели сложную структуру, каждый элемент которой может требовать специфических знаний из источников различной природы и применения различных технологий решения. Использование в этом случае жестких линейных схем, предлагаемых, например, Rational Rose, дает частичные решения, так как они рассчитаны на работу с одним типом источников – реляционными БД [4]. Кроме того, гипертрофированное развитие одних элементов технологий (например, KDD-методов) при отставании других (методов публикации результатов работы KDD) привело к потере цельного взгляда на разработку СППР и использованию частичных не всегда обоснованных решений в качестве целевых. Ситуация усугубляется тем, что некоторые задачи требуют знаний различной природы, которые размещены в сотнях территориально распределенных источников [9].

Совокупность перечисленных факторов привела к тому, что было нарушено главное требование к разработке СППР — релевантность целевого программного продукта семантике исходной задачи. Решение этой проблемы требует поиска новых технологий, которые обеспечат последовательную эволюцию исходной задачи в непротиворечивую СППР в рамках контекста релевантной ПО [3, 5].

В данной работе предложена технология построения релевантных СППР на основе знаний, полученных из распределенных гетерогенных источников. Методологической основой

технологии послужил алгоритм пошаговой детализации процесса решения через уточнение свойств ресурсов, задействованных на каждом этапе разработки СППР. В качестве ресурсов рассматриваются доступные источники знаний и готовые информационно-программные компоненты, что обеспечивает минимизацию затрат на решение исходной задачи.

1. Задача построения СППР

Постановку общей задачи сформулируем следующим образом. Имеется ЗПР Task, решение которой основано на знаниях ПО Knowledge. Для формирования Knowledge имеются ресурсы: Sources – источники знаний, Methods – методы добычи знаний, Tools – технологии реализации методов. Требуется разработать метатехнологию (далее технологию) использования ресурсов, обеспечивающую построение СППР, семантически и технологически релевантной Task. Основное требование к решению – поддержка ресурсов, реально используемых для построения СППР в настоящее время.

Рассмотрим вариант решения этой задачи для ЗПР, относящихся к группе decision making [2, 7]. Ситуацию принятия решений S в этом случае можно описать кортежем

где Master – лицо, принимающее решение (ЛПР); Task – постановка задачи; Goal – цель решения; Knowledge – Π O, содержащая альтернативные варианты решения; M – механизм выбора из Π O лучшего варианта достижения цели.

Практическая реализация модели S требует уточнения параметров, отражающих специфику поставленной задачи и максимально влияющих на качество целевой системы. В данном случае уточнение производится за счет источников знаний, методов и средств добычи знаний:

где Sources – источник получения знаний; Methods – методы добычи знаний, построения ΠO и принятия решения; Tools – технология реализации методов и представления решения; W – ограничения на решение (в основном это требование минимизации затрат).

Задача заключается в поиске значений параметров Knowledge, Sources, Methods, Tools, последовательное сочетание которых в процессе разработки СППР дает результат, релевантный задаче Task.

2. Концептуальная схема решения

Основное свойство модели S' – адаптивность, которая проявляется в возможности участников системы накапливать информацию и корректировать свое поведение с учетом повышения информированности за счет наблюдаемой истории их взаимодействия. Согласно логике задачи отправной точкой процесса адаптации является ЛПР, способное определить существенные свойства исходной задачи. На основании этой гипотезы построим стратегию реализации модели S', которую можно рассматривать как концептуальное решение поставленной задачи (рис.1).



Рис.1. Концептуальная схема построения СППР

Существенно, что в этой схеме роль ЛПР заключается в выборе управления, которое привело бы систему в наиболее предпочтительное, с его точки зрения, состояние.

Поиск релевантных сочетаний значений параметров модели S' предполагает наличие классификаций, ранжирующих параметры Task, Knowledge, Sources, Methods, Tools и лежащих в основе процесса физической реализации концептуального решения.

3. Классификация параметров модели

В литературе предлагаются различные варианты классификаций параметров СППР, однако единого варианта в настоящее время не существует. Авторы предлагают классификацию параметров модели S' на макроуровне, что расширяет спектр ее практического использования и оставляет возможность неограниченного уточнения.

3.1. Классификация ЗПР по значимости результатов решения

ЗПР принято классифицировать в зависимости от специфики задач и степени их структуризации, однако реалии конца XX – начала XXI вв. говорят о первоочередной необходимости классификации ЗПР по значимости и масштабам влияния результата решения на объект управления (ОУ). В качестве ОУ можно рассматривать государство, группу государств, предприятие и т. д. По этому признаку все СППР можно разделить на три группы:

- G1 задачи, которые не вызывают непоправимых последствий для OУ. Это задачи справочного, познавательного, обучающего характера.
- G2 задачи, которые могут иметь достаточно серьезные, но поправимые последствия локального масштаба. К ним относятся задачи управления любыми локальными процессами, где нарушение одного звена не влияет на жизнеспособность ОУ в целом.
- G3 задачи, которые вызывают широкомасштабные непоправимые последствия. К ним относятся задачи в области военного дела, медицины, управления сложными технологическими комплексами, неправильное решение которых несовместимо с жизнедеятельностью ОУ.

В данном случае ЗПР подразделяются на три группы, в общем случае их количество зависит от контекста задачи.

Оценив исходную задачу, необходимо оценить качество знаний, которые будут использованы при ее решении.

3.2. Классификация знаний по качеству

Резонно предположить, что различные ЗПР изначально требуют знаний различного качества. Ниже предлагается один из вариантов классификации знаний, соответствующий современному состоянию информационной индустрии.

Raw knowledge (RK) – потенциально полезные «сырые» знания, требующие интеллектуального обогащения (обработки человеком или «очистки» с помощью специальных методов и технологий) и пригодные для решения ЗПР с низкими требованиями к качеству результата. RK-знания часто избыточны, слабо структурированы, получены из различных, иногда анонимных, источников без подтвержденной юридической или какой-либо другой ответственности. В большинстве случаев RK-знания носят вспомогательный характер. Источники RK-знаний отличаются низким уровнем компетентности и используются для задач группы G1.

Half-knowledge (HK) – знания, обогащенные до уровня, достаточного для решения задач с четко определенными требованиями к качеству решения. В большинстве случаев НК-знания представляют собой результат аналитической обработки RK-знаний или знания, полученные из источников с повышенным уровнем компетентности. Применяются при решении задач G1.

Fidelity knowledge (FK) – знания, обладающие свойствами точности и полноты, достаточно выраженными для качественного решения поставленной задачи. Они адекватны решаемой задаче, получены из компетентных источников, их качество известно и подтверждено юридически. Применяются при решении задач G2 и G3.

Modern knowledge (МК) – точные и полные знания, обладающие, по сравнению с FK-классом, свойством актуальности (новизны). Такие знания имеют высокую стоимость и представляют наибольший интерес для компаний, деятельность которых основана на постоянном

внедрении новейших достижений науки и техники в производство. Применяются при решении задач G2 и G3.

После сопоставления задач и знаний рассмотрим источники их получения.

3.3. Классификация источников знаний

Источники знаний, присущие современному этапу развития постиндустриального общества, можно подразделить на три группы:

- базы данных;
- ресурсы Интернета;
- эксперты (лучшие в своей области знаний специалисты).

Базы данных (data base) и сформированные на их основе хранилища данных (data warehouse), а также витрины (киоски) данных (data marts) являются основным источником знаний для большинства существующих СППР, решающих задачи G2 и G3. Несмотря на различия в подходах и реализациях, всем СППР на основе БД свойственны следующие общие черты:

- предметная ориентированность;
- интегрированность;
- привязка ко времени;
- замкнутость;
- неизменяемость;
- компетентность;
- точность, полнота знаний в рамках ПО СППР.

Ресурсы Интернета (открытая зона) являются источником, компетентность которого трудно или невозможно оценить. Используются для решения задач G1, в задачах G2 и G3 могут быть задействованы зоны парольного доступа, но и в этом случае возникают проблемы по причине периодического несанкционированного проникновения даже в самые защищенные области Интернета, сопровождаемые ухудшением качества знаний.

Ограничения в применении вызваны следующими свойствами, которыми обладает в настоящее время сеть Интернет:

- анонимность источников информации;
- манипулирование навигацией пользователей;
- манипулирование сообществами пользователей;
- манипулирование предпочтениями пользователей и их потребностями.

Эти свойства сохранятся, вероятно, до тех пор, пока сеть Интернет не будет перестроена на основе Грид-технологий [11].

Эксперты в качестве источников знаний для решения ЗПР используются с тех пор, как такие задачи появились. Важнейшими преимуществами экспертных источников являются:

- точность;
- полнота;
- актуальность;
- высший уровень компетенции, юридическое подтверждение качества знаний.

Эти свойства позволяют рассматривать экспертов в качестве основного источника решения задач G3. Случай, когда эксперт получает знания из БД, можно считать частным, так как на эти знания, как правило, распространяется компетенция эксперта. То же самое можно сказать о закрытых областях Интернета, содержащих информацию.

Для добычи знаний из перечисленных источников применяются специализированные методы и средства, которые являются предметом исследований инженерии знаний.

3.4. Классификация методов и средств добычи знаний

Решение ЗПР на основе БД представляет собой технологически сложную задачу, решение которой требует применения специальных методик и инструментария. Современные решения основаны на CASE-технологиях (Computer-Aided Software/System Engineering) и инструментальных CASE-средствах, позволяющих максимально систематизировать и автоматизировать все этапы разработки СППР [5, 6]. В качестве примера можно привести CASE-средство

ERwin фирмы PLATINUM technology, S-Designer фирмы Sybase, Rational Rose компании Rational Software (рис. 2), использующие стандарты методологии IDEFO, IDEF3, IDEF1X [4, 6].

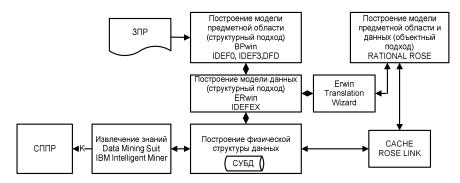


Рис. 2. Схема построения СППР на основе БД

На рис. 2 символом К обозначены так называемые «вспомогательные компоненты», в данном случае это средства консолидации знаний и публикации искомой ПО. В настоящее время К-средства являются самым слабым звеном технологий построения СППР.

Для извлечения знаний из БД используются две группы методов: «Data Mining» (DM) и «Knowledge Discovery in Databases» (KDD). DM-методы выполняют ограниченный спектр операций по поиску полезной информации – так называемых «паттернов» знаний из БД, для чего используется широкий спектр специализированных алгоритмов: A1, CN2, C4.5, DREV, ID3, OC1, Ripper [3–5]. KDD-методы применяют «паттерны» в качестве исходного материала для извлечения «чистых знаний». В качестве методической основы используются нейронные сети, деревья решений, эволюционное программирование, генетические алгоритмы, реализованные в ряде систем, наиболее развитыми из которых можно считать пакеты Data Mining Suit, IBM Intelligent Miner, Microsoft OLE DB for Data Mining.

Независимо от методов и средств реализации системы, созданные на основе БД, обладают следующими общими свойствами:

- хорошо формализованная предметная область;
- постоянные источники знаний локальной природы (реляционные БД);
- жесткая программная связь с постоянным количеством локальных источников (БД) на основе заранее известных форматов;
 - недостаточное развитие средств консолидации фрагментов в однородную ПО;
 - отсутствие независимых средств автоматизации публикации ПО;
 - сложность модификации ПО и компонентов СППР.

Решение ЗПР на основе ресурсов Интернета в настоящее время проходит стадию становления. Для добычи знаний применяются методы группы «intellectual search» (IS), реализованные в поисковых системах Google, Yahoo, IceRocket, Rambler и др. [5]. Для доступа к «поисковикам» используют web-браузеры MS Explorer, Opera, Netscape и др. Как правило, IS-методы поставляют «сырые» данные, являющиеся исходным материалом для описанных выше DM- и KDD-методов (рис. 3).

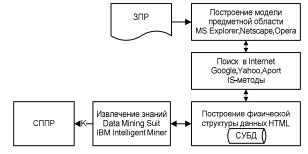


Рис. 3. Схема построения СППР на основе ресурсов Интернета

Несмотря на явные преимущества работы с территориально распределенными источниками и наличие средств фильтрации полученной информации, использование данной схемы для решения ЗПР ограничено перечисленными выше свойствами Интернета.

СППР на основе экспертных знаний, известные как экспертные системы (ЭС), появились в начале 1980-х гг. для решения задач диагностики. В создании ЭС участвует очная группа разработчиков, состоящая из эксперта (группы экспертов), инженера знаний и программиста. В разработке используются специализированные методы добычи экспертных знаний, формирования ПО и проведения диагностики.

В качестве методической основы извлечения имплицитных знаний используются различные методики проведения интервью. Формализация знаний осуществляется с помощью различных моделей, включая продукции, фреймы, семантические сети, и языков представления знаний OPS5, FRL, NET6 и др. [11]. Реализация моделей выполняется в рамках специализированных инструментальных средств, среди которых можно выделить TMYCIN, CLIPS, JESS, GENERIC, G2 [6]. Серьезные недостатки известных технологий данной группы: необходимость создания очной группы разработчиков, замкнутость архитектуры, сложность «внешнего» использования отдельных компонентов, ориентация на очную работу с экспертом – делают крайне сложной интеграцию готовых решений во внешние конструкции, в этом плане более интересны технологии работы с декларативными эксплицитными знаниями. Одна из них показана на рис. 4 и подробно описана в [8]. Основные преимущества данной технологии: отсутствие необходимости в очной группе разработчиков за счет высокой степени автоматизации всех этапов разработки СППР; возможность заочной работы с удаленными экспертами; универсальность применения компонента SPDprint, что позволяет использовать его для публикации ПО, нормализованных до уровня SGML-, HTML-, DOC-форматов (рис.4).

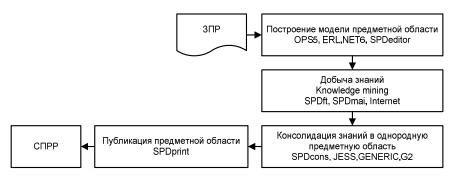


Рис. 4. Схема построения СППР на основе экспертных знаний

Применение данной схемы ограничено требованием предварительной договоренности с экспертами и их согласием на передачу знаний в дистанционном (заочном) режиме.

4. Стратегия применения

Реализация модели S' на основе предложенной системы классификаций позволяет уже на стадии проектирования:

- оценить исходную задачу;
- определить качество знаний, необходимых для ее решения;
- выбрать источники знаний, соответствующие уровню качества;
- выбрать методы добычи знаний, соответствующие источникам;
- выбрать программные компоненты, реализующие методы добычи;
- добыть знания и сформировать ПО;
- выбрать способ публикации ПО (локальный или сетевой);
- опубликовать ПО в форме СППР.

Применение указанных выше схем по отдельности особых проблем не вызывает. Однако современные задачи сложны и сильно структурированы, поэтому для их решения предлагаются новые подходы, включая «понятийные сети» (knowledge networks) [7]. Данный подход предпо-

лагает для каждой новой ЗПР создание временной сети источников и добычу FK-, МК-знаний для ее решения. В нашем случае схема организации «понятийной сети» будет включать все приведенные выше схемы решения (рис. 5).

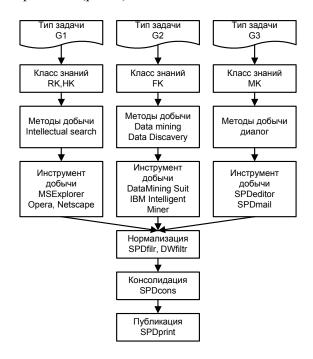


Рис. 5. Схема построения СПП на основе разнородных знаний

Выбрать по данной схеме траекторию решения, соответствующую природе и контексту исходной задачи, достаточно просто путем согласования параметров модели S' на базе приведенной выше классификации. Первые четыре этапа каждой ветви схемы определяются спецификой задачи и носят индивидуальный характер. Три последних этапа являются общими для всех ветвей схемы и служат средством консолидации и публикации ПО.

Предложенные решения содержат программные компоненты, которые в ряде случаев носят общий характер при незначительных ограничениях [8]. При появлении новых типов задач, методов и средств решения данную схему несложно откорректировать, адаптировав тем самым процесс разработки СППР к изменяющимся внешним условиям.

5. Пример использования

В качестве примера использования предложенной схемы построения СППР рассмотрим одну из актуальных задач, которую приходится периодически решать многим современным компаниям.

Предположим, IT-компания заключила крупный контракт на решение задачи Т на языке С#. Имеются текст постановки задачи Т, тексты подзадач Т1, Т2, Т3, Т4, Т5 и спецификация требований S(T1), S(T2), S(T3), S(T4), S(T5) к решению подзадач. Требуется разработать СППР с целью аккумуляции имеющихся в компании внутренних и внешних знаний в ПО, релевантную Т и доступную для использования территориально распределенному персоналу.

Определим типичные для данной задачи начальные условия. Пусть знания для решения подзадач T1, T2 имеются в БД компании. Новейшие технологии для решения задач T3, T4, T5 отсутствуют, но ими обладают внешние источники (эксперты), которые представлены в Интернете и априори достаточно компетентны.

Требование к решению: минимизация времени и средств на разработку, использование готовых программных компонентов.

Общая задача включает решение следующих подзадач:

1) добыча знаний из БД для решения Т1 и Т2;

- 2) добыча справочной информации о внешних источниках;
- 3) добыча экспертных знаний из внешних источников для решения Т3, Т4, Т5;
- 4) консолидация знаний, полученных из различных источников в однородную ПО;
- 5) публикация ПО в форме СППР.

Задача 1 относится к группе G2, источник знаний явно указан в постановке задачи:

 $G2 \rightarrow FK \rightarrow$ база данных $\rightarrow DM$ -методы $\rightarrow DataMining Suit <math>\rightarrow$ знания $\rightarrow SPBnorm \rightarrow Part1$.

Извлеченные из БД знания размещаются в файле Part1.

Задача 2 носит вспомогательный характер, относится к группе G1 и требует соответствующего решения:

 $G1 \rightarrow RK \rightarrow Интернет \rightarrow IS$ -методы $\rightarrow Explorer \rightarrow$ знания $\rightarrow \phi$ ильтрация \rightarrow эксперт.

Операция фильтрации (выбор конкретного эксперта из найденного списка, уточнение и т. д.) в данном случае выполняется человеком.

Задача 3 касается инноваций, жизненно важна для компании, относится к группе G3 и требует соответствующих этой группе этапов решения:

$$G3 \rightarrow MK \rightarrow SPDeditor \rightarrow SPDmail \rightarrow знания \rightarrow SPDnorm \rightarrow Part2.$$

Полученная от экспертов информация размещается в файле Part2.

Решение задач 4 и 5 носит формальный характер и заключается в объединении двух фрагментов в однородную ПО и публикации результата:

$$SPDkons(Part1, Part2) \rightarrow SPDprint \rightarrow C\Pi\Pi P.$$

Таким образом, исходная задача решена, причем использование средств автоматической публикации ПО SPDprint позволило минимизировать стоимость разработки и сделать целевую систему доступной для территориально распределенного персонала.

Описанные схемы работы с фрагментарными знаниями применялись, в частности, при разработке системы «Наука Республики Беларусь», где источником цифровой информации (статистической отчетности по науке) служили БД, текстовая информация (аналитические обзоры) доставлялась из интернет-ресурсов, а мультимедиа-презентации научных организаций пересылали по электронной почте удаленные эксперты. Полученные результаты — рабочие версии системы за 2002 и 2003 гг. — говорят о полезности данного подхода. Значимый эффект был получен, в частности, за счет применения в компоненте консолидации SPDkons оригинальных фильтров очистки, что по сравнению с функцией WORD save as web позволило на 25-30% снизить информационный шум и соответственно уменьшить физический размер файлов предметных областей.

Заключение

В статье рассмотрены проблемы построения СППР на основе знаний, полученных из распределенных гетерогенных источников. Показано, что сложность и масштаб ЗПР увеличиваются, изменяются условия их решения и природа источников знаний. Обоснована необходимость поиска новых схем построения СППР, в большей степени учитывающих динамику изменения внешней среды. В основу моделирования знаний положен онтологический подход, программная часть реализует корректную композицию программных и информационных компонентов.

Получены следующие основные результаты:

- построена модель СППР, выделены ее активные составляющие, максимально влияющие на качество принятия решений;
- введена система классификации активных составляющих, которая позволяет сопоставить семантику задачи, качество знаний, методы решения и средства их реализации;
- предложена стратегия построения СППР, которая позволяет оценить исходную задачу и качество необходимых для решения знаний, выбрать соответствующий метод решения и технологию его реализации, а также оформить результат в форме СППР;

– представлен пример использования нового подхода для решения типовой задачи. Полученные результаты апробированы в процессе разработки ряда прикладных систем, нашедших достаточно широкое применение на практике, в частности в системах «Forensic Pathology» и «Orthopedische Casuïstiek» [10].

За рамками статьи остались вопросы построения СППР на основе корректной композиции БД и готовых DDL-модулей. Решение этой проблемы требует разработки промежуточного слоя для согласования требований исходной задачи, спецификаций БД и программных компонентов, что представляет собой отдельную технологически сложную задачу.

Список литературы

- 1. Сидоров И.И. Логистическая концепция управления предприятием. СПб.: Знание, 2001. 168 с.
- 2. Железко Б.А. Системы поддержки принятия решений: вопросы создания и примеры использования. Мн.: КИВТ НАН Беларуси, 1998. 80 с.
- 3. Goossenaerts J., Pelletier C. Enterprise ontologies and knowledge management // Proceedings of the 7th International Conference on Concurrent Enterprising «Engineering the Knowledge Economy through Co-operation». Bremen, Germany, 2001. P. 281-285.
- 4. Саймон А.Р. Стратегические технологии баз данных: менеджмент на 2000 год /Под ред. М.Р. Когаловского. М.: Финансы и статистика, 1999. 479 с.
- 5. Han J., Kamber M. Data Mining. Concept and Techniques. Morgan Kaufman Publishers, 2000. 431 p.
- 6. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.
- 7. Holsapple C.W. Knowledge management Support of Decision Making // Decision Support Systems. 2001. № 31. P.1-3.
- 8. Краснопрошин В.В., Шаках Γ ., Вальвачев А.Н. Интеграция распределенных экспертных знаний: проблемы и решения // Информатика. -2004. -№ 1. -C. 45-52.
- 9. Krasnoproshin V., Obraztsov V., Vissia H. Decision-Making by Precedence: Modeling, Technology and Applications // Proceedings of International Conference on Modeling and Simulation in Technical and Social Sciences (MS'2002). Girona, Spain, 25-27 June 2002. P. 267-277.
- 10. Krasnoproshin V., Valvachev A., Vissia H. Unstructured Knowledge Synthesis for Decision-Making Problems // Proceedings of the Seventh International Conference PRIP'2003. Minsk. V.1. 2003. P. 145-149.
- 11. Foster I., Kesselman C. The Grid: Blueprint for a New Computing Infrastructure. San-Francisco: Morgan-Kaufmann Pub, 2004. 462 p.

Поступила 13.08.04

Белорусский государственный университет Минск, пр. Ф.Скорины, 4 e-mail: kras@bsu.by

V.V. Krasnoproshin, G. Shakah, A.N. Valvachev

TECHNOLOGY OF BUILDING DECISION SUPPORT SYSTEMS BASED ON DISTRIBUTED GETEROGENEOUS KNOWLEDGE

The paper deals with the problems of decision support systems construction based on heterogeneous knowledge obtained from different sources. A model of a system and strategy of its realization to obtain the result which is relevant to the sources problem are proposed. Special attention is paid to distinctive features of software technologies generally recognized in knowledge extraction.