

ческой диагностики / И. В. Ковалев, А. А. Ступина, Р. Ю. Царев // Приборы и системы. Управление, контроль, диагностика. 2006. № 12. С. 47–52.

3. Бозм, Б. Характеристики качества программного обеспечения / Б. Бозм, Дж. Браун, Х. Каспар и др. М. : Мир, 1999.

4. Глас, Р. Руководство по надежному программированию : пер. с англ. / Р. Глас. М. : Финансы и статистика, 2000.

5. Дилон, Б. Инженерные методы обеспечения надежности систем / Б. Дилон, И. Сингх. М. : Мир, 1992.

6. Майерс, Г. Надежность программного обеспечения : пер. с англ. / Г. Майерс ; под ред. В. Ш. Кауфмана. М. : Мир, 1998.

7. Мамиконов, А. Г. Типизация разработки модульных систем обработки данных / А. Г. Мамиконов, В. В. Кульба, С. А. Косяченко. М. : Наука, 1989.

8. Мамиконов, А. Г. Синтез оптимальных модульных систем обработки данных / А. Г. Мамиконов, В. В. Кульба. М. : Наука, 1986.

9. Орлов, С. А. Технологии разработки программного обеспечения / С. А. Орлов. СПб. : Питер, 2002.

O. A. Antamoshkin, Ju. A. Nurgaleeva, A. V. Usachev

MODULAR DECOMPOSITION MODELS OF TECHNOLOGICAL CONTROL CYCLES SOFTWARE

Modular decomposition models of technological control cycles software are considered.

Keywords: software, technological control cycles, modular decomposition.

© Антамошкин О. А., Нургалева Ю. А., Усачев А. В., 2009

УДК 681.3

П. В. Зеленков, М. А. Селиванова, В. В. Брезницкая, А. П. Хохлов

МОДУЛЬ ОБРАБОТКИ ИНФОРМАЦИОННЫХ ЗАПРОСОВ ПОЛЬЗОВАТЕЛЕЙ В СЕТЬ ИНТЕРНЕТ ДЛЯ КОРПОРАТИВНЫХ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИХ СИСТЕМ

Предложена новая технология организации запросов пользователей в сеть Интернет, применяемая в корпоративных информационно-управляющих системах. Предложена структура модуля обработки запросов пользователя, состоящая из агентов: получения запросов, распознавания профиля пользователя, распознавания запроса, сравнения типовых запросов, поиска информации и отображения отклика системы. Применение предлагаемой технологии позволяет существенно снизить внешний интернет-трафик.

Ключевые слова: корпоративные информационно-управляющие системы, Интернет, пользователь сети, структура системы, пользовательский запрос.

В настоящее время большинство компаний для информационного обеспечения сотрудников при взаимодействии с внешними информационными ресурсами используют подключение к глобальной сети Интернет. Для этого применяется наиболее популярная схема подключения (рис. 1). Данная схема является очень удобной, что связано с возможностью контролировать действия сотрудников в глобальной сети и закрывать доступ к сомнительным сайтам, таким как набирающие популярность социальные сети, развлекательные порталы и т. п.

В классической схеме подключения (см. рис. 1) пользователь может напрямую обратиться к Proxy-серверу и работать в сети Интернет без использования корпоративной информационно-управляющей системы (КИУС). Это приводит к тому, что запросы пользователя, связанные с рабочим процессом, никак не учитываются в рамках этой системы. Если рассмотреть работу

сотрудников одного отдела в рамках предприятия, то можно увидеть, что они, как правило, работают с одной и той же информацией. Кроме того, большинство сотрудников при входе в сеть Интернет просматривают одни и те же информационные и новостные ресурсы. Следовательно, интернет-трафик, затрачиваемый на просмотр данных ресурсов, расходуется нерационально: он увеличивается соответственно числу сотрудников или более, если один сотрудник просматривает данный ресурс неоднократно.

Необходимо отметить, что в небольших компаниях выход в сеть Интернет может быть организован вообще напрямую, т. е. без использования Proxy-сервера. Это влечет за собой еще более нерациональное использование интернет-ресурсов и увеличение затрат рабочего времени сотрудников, не связанных с выполнением рабочих процессов.

Для решения указанных выше проблем предлагается организовать обработку запросов пользователей через единый модуль в КИУС (рис. 2).

В качестве подобного модуля может выступать единый корпоративный информационно-управляющий браузер. Таким образом, все запросы пользователей будут проходить через данный модуль. Основная задача модуля при этом состоит в том, чтобы собирать предпочтения пользователей информационной системы, распознавать пользователей и работать с ними с применением технологии персонификации, организовывать внутреннее хранения результатов отклика Интернета. Кроме того, для снижения загруженности интернет-трафика и его экономии на этот модуль возлагается организация специальной поисковой процедуры. Ее основная функция состоит в том, чтобы сравнивать запросы пользователей, объединяя пользователей в группы в соответствии с их предпочтениями и производить поиск информации сначала в локальной корпоративной сети, а уже потом в сети Интернет.

Предлагаемая архитектура модуля обработки пользовательских запросов основана на агентной технологии (рис. 3). В структуру модуля входит семь агентов:

- получения запросов;
- распознавания пользователя;
- распознавания запроса;
- сравнения типовых запросов;
- поиска информации в КИУС;
- поиска информации в Интернете;
- отображения отклика системы.

Рассмотрим назначение каждого агента более детально.

Агент получения запросов отвечает за организацию взаимосвязи пользователя корпоративной системы и КИУС при

проведении поисковой процедуры. Этот агент непосредственно принимает пользовательский запрос и передает его в систему. В результате работы данного агента запускается группа других агентов: распознавания пользователя, распознавания запроса, сравнения типовых запросов.

Агент распознавания пользователя организует персонифицированную работу системы с каждым отдельным пользователем. Его основная задача – это управление профилями пользователя, отслеживание его информационных потребностей как в краткосрочном, так и долгосрочном временном интервале. Очевидно, что сформированный профиль пользователя КИУС отражается на выборе рациональной стратегии проведения информационно-поисковой процедуры.

Выбор категории или профиля пользователя нередко определяется набором типовых информационных потребностей. При формировании профиля пользователя часто учитывается ряд атрибутов, характеризующих различные производственные информационные потребности, возникающие в процессе функционирования предприятия или организации. Можно отметить наиболее распространенные атрибуты:

- наименование информационных разделов КИУС, которые необходимы пользователю для выполнения должностных обязанностей. Этот атрибут позволяет учитывать группы профилей, исходя из должностной принадлежности каждого отдельного пользователя, а также сформировать группы пользователей, исходя из других информационных потребностей (краткосрочных и периодических рабочих информационных потребностей, хобби т. п.);
- отслеживание предпочтений как на уровне группы, так и на уровне конкретного пользователя в том или ином

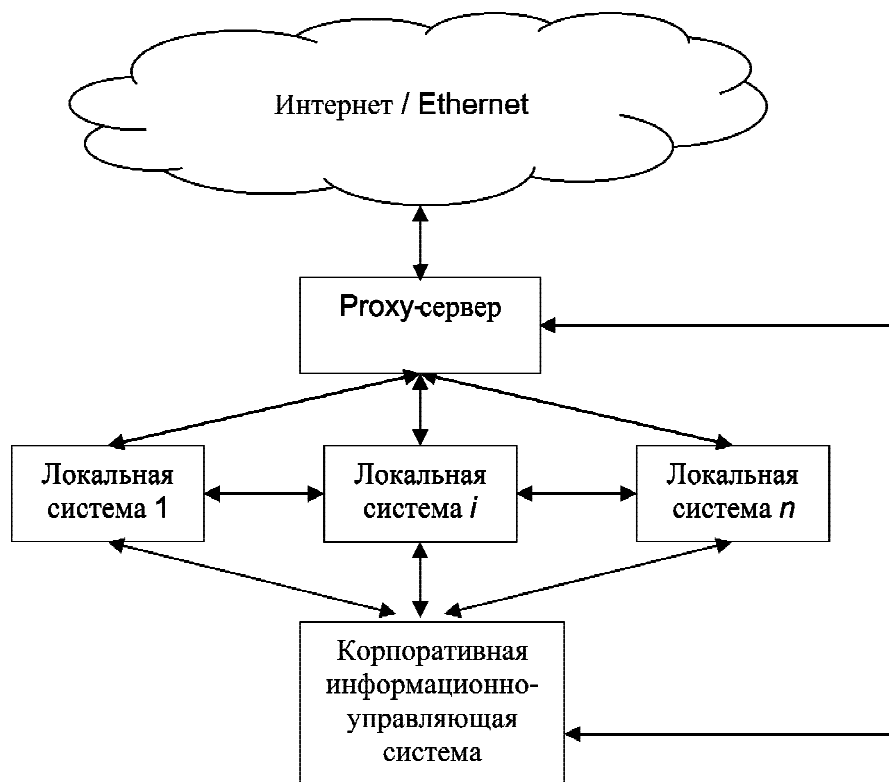


Рис. 1. Классическая схема организации поисковой процедуры в корпоративной информационно-управляющей системе

информационном разделе или ресурсе. Данный атрибут позволяет динамически менять краткосрочный и долгосрочный профиль пользователя;

– история проведенных пользователем и группой пользователей поисковых процедур. Этот атрибут отве-

чает за более качественное ранжирование результатов проведения поисковых процедур;

– динамика изменения пользовательских информационных потребностей. Данный атрибут позволяет организовать устоявшиеся группы пользователей и, следователь-

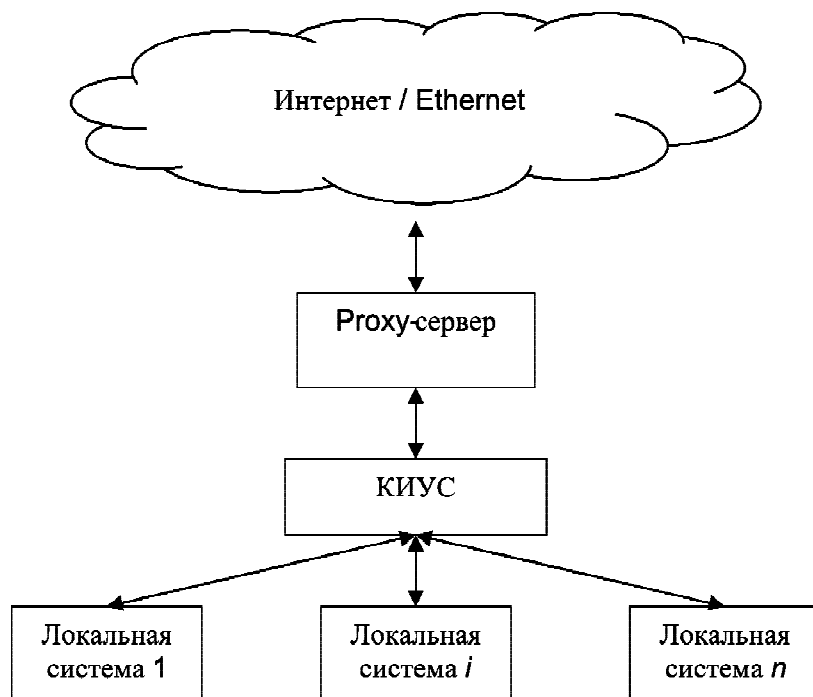


Рис. 2. Предлагаемая структура организации информационно-поисковой процедуры в КИУС

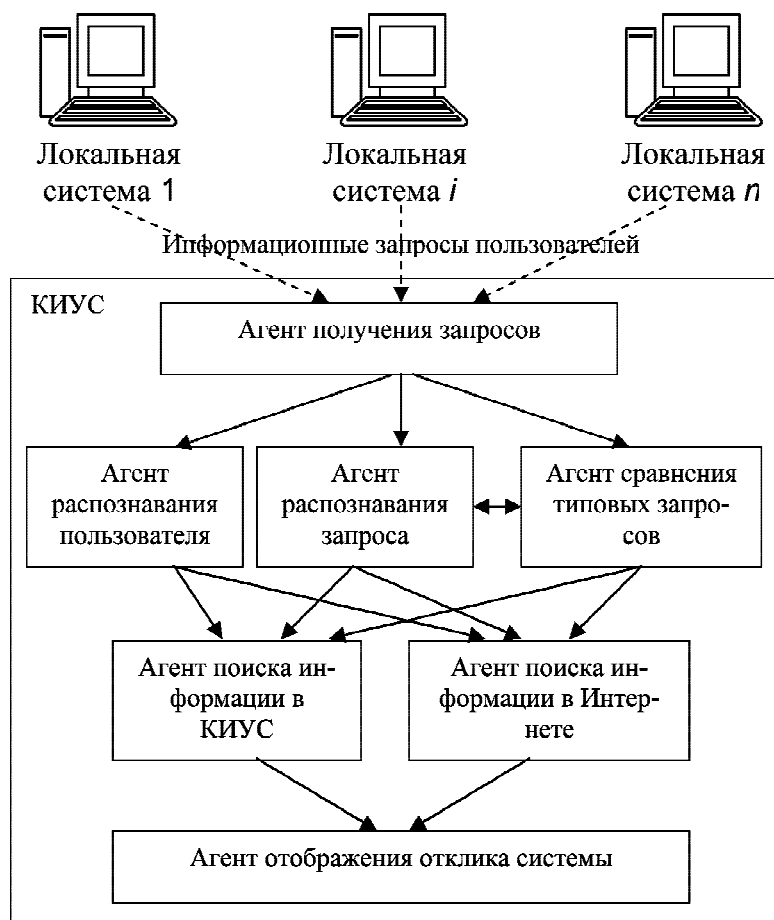


Рис. 3. Структура модуля обработки информационных запросов пользователя

но, более качественно управлять их информационными потребностями.

Агент распознавания запроса отвечает за корректный прием пользовательского запроса и его подготовку к отправке в информационное пространство КИУС. В это информационное пространство входят информационные ресурсы компании и интернет-ресурсы, к которым организован доступ сотрудников предприятия.

Агент сравнения типовых запросов связан с поиском в проиндексированной базе уже проведенных поисковых процедур по заданному пользователем запросу. Он выбирает группы пользователей, которые проводили данную ранее поисковую процедуру и проверяет категорию полученного информационного ресурса. В соответствии с данной информацией происходит коррекция параметров ранжирования информации дальнейшими агентами рассматриваемого модуля обработки запросов.

Агент поиска информации в КИУС отвечает за проведение поисковой процедуры во внутренних информационных ресурсах.

Повысить качество проведения этой процедуры можно за счет реализации функции каталогизации информации. Для этого необходимо реализовать в рамках КИУС проверку релевантности информации по заданным информационным категориям. Для этого можно воспользоваться технологией анализа информации, основанной на применении частотных словарей и тезаурусов [1; 2]. Таким образом, вся информация в корпоративной системе будет поделена на категории, которые будут соответствовать информационным разделам, входящим в профиль пользователей.

Агент поиска информации в Интернете отвечает за проведение поисковой процедуры по заданным параметрам.

Необходимо отметить, что на этом этапе применяется технология метапоиска [3; 4].

Метапоисковая система – это система, которая предоставляет единый доступ к нескольким другим поисковым системам, т. е. обслуживает запросы пользователей за счет опрашивания других поисковых систем. Современные метапоисковые системы имеют встроенные механизмы ранжирования и проверки релевантности информации. Такие системы популярны в силу следующих причин:

- повышения полноты охвата. Необходимо помнить, что ни одна поисковая система не имеет полного покрытия в Интернете, а использование нескольких поисковых систем повышает вероятность обнаружения искомого документа (рис. 4). Например, полностью русскоязычный сегмент сети не будет проиндексирован в базах данных поисковых систем, также есть документы, которые встречаются во всех трех приведенных системах, а есть сектора, которые учтены только в одной системе;

- повышения качества поиска. Наиболее типичный способ – выбор тех поисковых систем, которые более всего соответствуют текущим потребностям пользователя. Например, это может быть использование специализированной поисковой системы вместо поисковой системы общего назначения;

- новых возможностей поиска. Крупномасштабная поисковая система, такая как Yandex или Google, не может тратить много времени на обработку каждого отдельного запроса из-за их огромного количества. Метапоисковые системы не имеют такого ограничения и могут фокусироваться на решении специализированных задач поиска, ориентированных на узкую целевую аудиторию. Все это облегчает внедрение новых методов поиска, по-

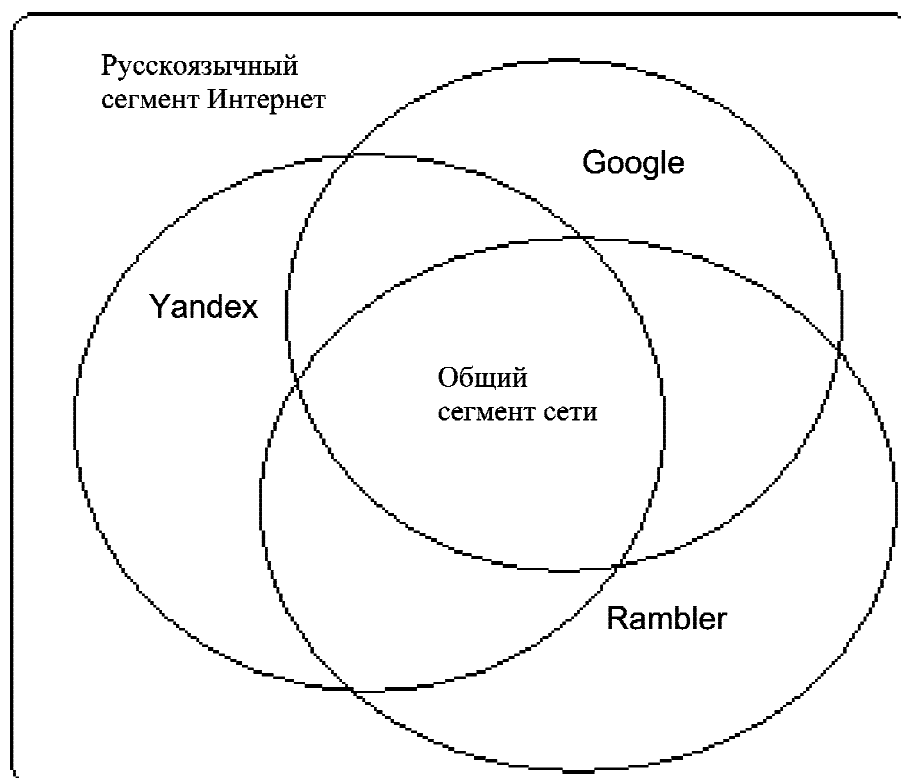


Рис. 4. Пример русскоязычного сегмента сети Интернет

скольку дает возможность проверить их эффективность без реализации полноценной поисковой системы.

При построении метапоисковых систем приходится решать ряд проблем. Так, языки запросов, используемые в разных поисковых системах, зачастую сильно отличаются, и поэтому необходим либо сильно упрощенный язык поиска для метапоисковой системы, либо модификация запросов для каждой конкретной поисковой системы. Другой проблемой является слияние ответов от различных поисковых систем. В этом вопросе исследователи поисковых систем пока не пришли к единому решению [3].

В основу работы агента отображения отклика системы заложены принципы обработки информации, полученной от всех предыдущих агентов. Вся информация делится на две группы: полученные информационные документы из корпоративной системы и из Интернета, а также информация о пользователе и его профиле. При анализе профиля выполняется анализ как краткосрочного профиля, так и долгосрочного, при этом также необходимо анализировать историю проведения поисковых процедур и самим пользователем, и группой (группами), к которой данный пользователь относится.

При отображении отклика системы очень важную роль играет механизм ранжирования полученной информации. Рассмотрим несколько подходов к ранжированию информации [4].

Алгоритм Клейнберга (HITS) основан на применении метода латентного семантического индексирования при ранжировании результатов, выдаваемых информационно-поисковыми системами, использующими принцип цитирования.

Алгоритм HITS обеспечивает выбор из информационного потока лучших авторов (первоисточников) и посредников (документов, от которых идут ссылки цитирования). Страница является хорошим посредником, если она содержит ссылки на ценные первоисточники, и наоборот, страница является хорошим первоисточником, если она упоминается хорошими посредниками.

Недостатки HITS следующие:

- отсутствие стабильности качества результатов HITS;
- алгоритм вычисления рангов HITS влечет рост рангов страниц при увеличении количества и степени связанности страниц соответствующего сообщества. В этом случае в результат может попасть много страниц на темы, не соответствующие информационной потребности пользователя, т. е., часть выдаваемых результатов, соответствующих требуемой теме, может оказаться не доминирующей. Это обуславливает присвоение высших рангов страницам на тему, не требуемую пользователем, т. е. происходит смещение тематики.

Алгоритм PageRank основан на том, что в отличие от литературного индекса цитирования не все ссылки считаются равнозначными. Этот алгоритм подсчитывает общий авторитет документа, в то время как алгоритм HITS определяет авторитет документа для конкретной темы.

Алгоритм PageRank был развит в 1996 г. в Стенфордском университете Л. Пейджем и С. Брином.

Этот алгоритм применяется в системе ранжирования, используемой в поисковой системе Google. Однако он

применяется не в чистом виде, а в модифицированном виде (Hilltop).

Недостатки алгоритма ранжирования Hilltop в Google следующие:

- он основывается на предположении, что каждый экспертный документ, который он находит, будет беспристрастен, свободен от спама и манипуляций;
- алгоритм старается выбирать страницы, за которые проголосуют как за авторитетные. Но нет никакой гарантии, что эти страницы также будут качественными;
- для его функционирования требуется значительная процессорная мощность;
- новые сайты или документы увеличивают сложность процесса ранжирования;
- поскольку большинство коммерческих сайтов весьма легко ссылаются на директории торговых ассоциаций, правительственные сайты, сайты образовательных учреждений, то такие сайты заполняют первую десятку по ранжированию на страницах результатов, что зачастую будет вести к некорректному результату.

В весовом алгоритме ранжирования текстовых сообщений ключевым словам из документов, выдаваемых информационно-поисковой системой, приписывается некоторый вес. Вес документа определяется как средний вес входящих в него значимых ключевых слов. Очевидно, что чем меньше этот вес, тем документ более уникален.

Ранжирование по Хиршу состоит в подсчете числа h публикаций одного автора, на которые имеется не менее h ссылок.

Этот метод был предложен Й. Хиршем в 2005 г. для оценки научных публикаций, чтобы обеспечить более высокую точность, и, что особенно важно, объективность по сравнению с получившим широкое распространение индексом цитирования.

Необходимо отметить, что разработчики алгоритмов ранжирования постоянно пытаются их совершенствовать, преследуя, как правило, две главные цели – улучшение качества поиска и уменьшение возможности искусственных воздействий на ранжирование результатов.

Таким образом, в статье дано описание предлагаемой авторами современной технологии организации запросов пользователей в сеть Интернет, которая может быть эффективно применена в КИУС. Данная технология позволяет повысить качество обработки запроса пользователя как в рамках корпоративного информационного пространства, так и при работе с Интернетом. Кроме того, эта технология существенно снижает внешний интернет-трафик за счет организации единой корпоративной браузерной системы с хранением и обработкой всей запрашиваемой пользователями информации.

Библиографический список

1. Мультилингвистическая технология поиска данных для подготовки и принятия решения в информационно-управляющих системах / П. В. Зеленков, П. М. Лохмаков, И. В. Ковалев, С. С. Огнерубов // Прогр. продукты и системы. 2007. № 2. С. 97–99.

2. Зеленков, П. В. Реализация модуля поиска информации на основе использования мультILINGвистических тезаурусов / П. В. Зеленков, М. В. Карасева // Вестник СибГАУ. 2009. Вып. 1 (22). С. 64–66.

3. Зеленков, П. В. Метапоисковая мультILINGвистическая система / П. В. Зеленков, М. В. Карасева,

И. Н. Карцан // Вестник СибГАУ. 2007. Вып. 3 (16). С. 69–70.

4. Зеленков, П. В. Модели и алгоритмы поиска и обработки информации в распределенных информационных системах : монография / П. В. Зеленков, И. В. Ковалев, В. В. Брезицкая. М. : Макс-пресс, 2008.

P. V. Zelenkov, M. A. Selivanova, V. V. Brezitskaya, A. P. Hoxhlov

INFORMATION REQUESTS PROCESSED MODULE FROM USERS IN THE INTERNET FOR CORPORATE INFORMATION SYSTEMS

A modern technology of the user requests organization to the Internet in the corporate information systems is proposed. A structure of the module processing user requests consisting of agents: receiving requests, recognition of the user profile, recognition of the user request, compare model requests, information retrieval and display system response was proposed. The application of the proposed technology allows you to reduce significantly the external Internet traffic.

Keywords: corporate information systems, Internet, user network, system structure, user request.

© Зеленков П. В., Селиванова М. А., Брезицкая В. В., Хохлов А. П., 2009

УДК 658.512.001.56

Т. А. Рукавицына

РАЗВИТИЕ МОДЕЛИ МЕТОДОЛОГИИ DEA

Рассмотрен метод DEA (Data Envelopment Analysis) и его модификация. Введено понятие границы отставания. Исследовано пересечение границы отставания и границы эффективности, получаемой по методу DEA. Построена гиперплоскость, проходящая через пересечение двух границ и делящая множество объектов на два класса.

Ключевые слова: DEA, эффективность, граница эффективности.

Проблема оценки эффективности функционирования предприятий очень остро встает в последние годы во многих сферах производства и сбыта продукции. Часто возникают задачи сравнения между собой и упорядочивания структурных подразделений и предприятий или организаций в целом по некоторому свойству, не поддающемуся непосредственному измерению [1]. При этом общее представление о степени проявления анализируемого латентного свойства складывается в результате определенного суммирования целого ряда частных (и поддающихся измерению) характеристик, от которых в конечном счете зависит это свойство. Несомненно, главным понятием здесь является понятие эффективности как наиболее общего, определяющего свойства любой целенаправленной деятельности, которое с познавательной точки зрения раскрывается через категорию цели и объективно выражается степенью достижения цели с учетом затрат ресурсов и времени. Поэтому оценка эффективности функционирования предприятий и организаций является очень важной для принятия верных управленческих решений.

В данной статье мы будем рассматривать метод DEA (Data Envelopment Analysis) основанный на построении

границы эффективности. Эта граница имеет форму выпуклой оболочки и используется в качестве эталона (точки отсчета) для получения численного значения оценки эффективности каждого из объектов в исследуемой совокупности.

По аналогии с границей эффективности можно ввести понятие границы отставания. Эта граница будет использоваться в качестве эталона для получения численного значения оценки отставания (убыточности) каждого из объектов. На пересечении двух границ будет построена гиперплоскость, разделяющая совокупность объектов на два класса.

Метод Data Envelopment Analysis. Данный метод представляет способ оценки производственной функции, которая неизвестна. Метод DEA основан на построении границы эффективности, которая и является аналогом производственной функции для случая, когда выпуск является не скалярным, а векторным, т. е. когда выпускается несколько видов продукции. Эта граница имеет форму выпуклой оболочки или выпуклого конуса в пространстве входных и выходных переменных, описывающих каждый объект в исследуемой совокупности. Из названия метода следует, что граница эффективности как бы опи-