

предметной области. Кроме того, в выбранном тексте релевантным предметной области может быть не весь документ, а только его часть, например отдельные разделы из учебников общего назначения, отдельные статьи из сборников статей, разделы из отчетов организаций и т. д. Учет этого ограничения поможет при принятии решения о предоставлении пользователю только необходимой ему части информации.

Следующий агент – это агент ранжирования. Он является не менее важным при обработке информации, так как при выдаче пользователю нескольких тысяч документов на первом месте в списке отображения должны стоять самые ценные документы.

*Агент тематико-ориентированного мониторинга* отвечает за анализ информационных предпочтений пользователя корпоративной системы в рамках информационно-тематических коллекций и обеспечение ему персонализированной поддержки навигации и персонализированных данных. Благодаря предоставлению пользователям информационных коллекций персонализированных навигационных меню из ссылок на страницы, близкие к их тематическим предпочтениям, происхо-

дит уменьшение времени, необходимого для поиска нужной информации, и снижается пользовательский трафик как в корпоративной, так и во внешней сети за счет просмотра только качественной информации.

Таким образом, предлагаемые решения должны повысить удобство работы пользователей с информационными ресурсами корпоративной системы и послужить для них дополнительным стимулом для более частого посещения этих ресурсов. Данный подход также должен существенным образом снизить нагрузку как на внутренний (корпоративный), так и на внешний трафик.

#### Библиографический список

1. Мультилингвистическая модель распределенной системы на основе тезауруса / П. В. Зеленков, С. В. Рогов, И. В. Ковалев, М. В. Карасева // Вестник СибГАУ. Красноярск, 2008. № 1 (18). С. 26.–28.
2. Зеленков, П. В. Метапоисковая мультилингвистическая система / П. В. Зеленков, И. Н. Карцан, М. В. Карасева // Вестник СибГАУ. Красноярск, 2007. Вып. 3 (16). С. 69–70.

P. V. Zelenkov, V. V. Brezitskaya, M. V. Karaseva, A. P. Khohlov

### PROCESSING SYSTEM OF THE HIGHLY TAILORED INFORMATION IN NETWORKS

*A new structure of the system formation and control by highly tailored information in networks is offered in the paper. The main peculiarity of this structure is that it assumes multilingual information processing within the bounds of one user's inquiry.*

*Keywords: multiagent system, meta-search, highly tailored information, network.*

УДК 378.016

Г. А. Доррер, А. А. Попов, К. В. Сысенко

### ИССЛЕДОВАНИЕ ЖИЗНЕННОГО ЦИКЛА ЭЛЕКТРОННЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ

*Предложена вероятностная модель жизненного цикла информационных ресурсов (ИР), основанная на формализме цепей Маркова. В качестве исходных предположений приняты следующие:*

– ИР классифицируются в соответствии с Дублинским ядром метаданных (*Dublin Core Metadata Element Set, DCMI*);

– выделено пять типовых бизнес-процессов по управлению ИР в соответствии с классификацией *American Productivity & Quality Center*;

– по важности информация в ИР классифицируется как критическая, важная и маловажная.

*Предложенная модель позволяет с помощью разработанной авторами программной системы оценить жизненный цикл всех компонентов единичного информационного ресурса, а также динамику ИР в организации при заданном законе поступления новых ресурсов. Это дает возможность объективно оценивать потребные объемы дискового пространства организации. Приведен пример моделирования годичной динамики ИР в одном из отделов администрации города.*

*Ключевые слова: информационные ресурсы, жизненный цикл, моделирование.*

В настоящее время информационные технологии (ИТ) стали неотъемлемой частью любой деятельности. В ряде случаев они играют обеспечивающую роль, в других слу-

чаях являются основой бизнеса. Сформировалось самостоятельное направление менеджмента – сервис-менеджмент информационных технологий (ITSM). Классичес-

ким руководством в данной области стала книга Я. Ниессена «Введение в ИТ Сервис-менеджмент» [1]. Предложенный в ней подход стал фактическим стандартом для формирования ИТ-служб предприятий как современных бизнес-подразделений, постоянно ориентированных на потребности своих пользователей. В дальнейшем такие структуры мы, следуя этой работе, будем называть ИТ-организациями.

Одним из процессов, реализуемых при менеджменте ИТ, является так называемое управление мощностями (*Capacity Management*). Задачей процесса управления мощностями является предоставление в нужное время и в экономически эффективной форме необходимых мощностей для обработки и хранения данных и обеспечение соответствующего баланса мощностей в ИТ-организации. Эффективное управление мощностями исключает панические закупки в последнюю минуту или покупку самой большой системы «на всякий случай», дорого обходящиеся ИТ-организациям. Многие центры обработки данных, например, постоянно работают с недогрузкой на 30...40 % или больше. Это не так плохо, если в ИТ-организации небольшое количество серверов. Но если имеются сотни и тысячи серверов, как у многих ИТ-организаций масштаба предприятия, то эти проценты означают потерю огромных финансовых средств [1].

В процессе управления мощностями важную роль играет управление информационными ресурсами (ИР), которые служат основой для создания информационных продуктов и оказания информационных услуг.

В широком смысле под информационным ресурсом понимаются «знания, подготовленные людьми для социального использования в обществе и зафиксированные на материальных носителях в виде документов, баз данных и знаний, алгоритмов, компьютерных программ, а также произведений искусства, литературы и науки» [2]. Информационные ресурсы организации, региона, страны рассматриваются как стратегические ресурсы, аналогичные по значимости запасам сырья, энергии и т. п.

Ценность информации, содержащейся в конкретном информационном ресурсе, может быть различной, кроме того, она изменяется во времени. Как и все на свете, информационный ресурс рождается, живет и умирает. Поэтому можно говорить о жизненном цикле информационного ресурса, в том числе и электронного. К сожалению, этот вопрос до сих пор недостаточно изучен в отличие, например, от жизненного цикла программного обеспечения и информационных систем [3]. Для управления жизненным циклом могут применяться технологии иерархического хранения данных HSM (*Hierarchical Storage Management*) [4], которые из разнородных средств хранения создают виртуальную файловую систему и управляют переносом данных с одного носителя на другой. Однако нам не известны математические модели жизненного цикла ИР, которые могли бы использоваться при разработке систем хранения данных.

Авторами предлагается сравнительно простая вероятностная модель жизненного цикла ИР, основанная на формализме цепей Маркова. Такая модель позволяет оценить жизненный цикл единичного ИР, а также динамику ИР в ИТ-организации при заданном законе поступления новых ресурсов.

**Модель жизненного цикла ИР.** Примем следующие предположения:

1. Информационные ресурсы классифицируются в соответствии с Дублинским ядром метаданных (*Dublin Core Metadata Element Set, DCMI*) [5]. DCMI представляет собой инвариантную к предметной области композицию наиболее общих полей описания ИР, введенную для обеспечения глобальной интероперабельности приложений, работающих с метаданными. Приведем краткое описание видов ресурсов:

- *коллекция* – множество, содержащее различные элементы;
- *данные* – информация, представленная в определенной структуре;
- *событие* – непродолжительное, ограниченное во времени явление;
- *изображение* – визуальное представление, отличное от текста;
- *сервис* – выполнение функций для конечного пользователя;
- *программные средства* – компьютерная программа;
- *аудио* – звуковое представление данных;
- *текст* – слова для чтения.

2. В соответствии с классификацией бизнес-процессов, разработанной Американским центром производительности и качества (*American Productivity & Quality Center*), выделим следующие бизнес-процессы по управлению ИР [6], которые мы рассматриваем как этапы жизненного цикла ИР:

- создание;
- хранение;
- обработка (упорядочение, поиск, изменение и обновление);
- архивирование;
- уничтожение.

3. В течение жизненного цикла меняется актуальность информации, содержащейся в информационном ресурсе. С этой точки зрения информация может классифицироваться как критическая, важная и маловажная [7]. Как правило, актуальность информации уменьшается с течением времени, следовательно эти классы актуальности наступают последовательно.

В дополнение к классификации [6] мы дополнительно ввели в жизненный цикл ИР этап архивирования, который обычно наступает, когда информация становится маловажной либо неактуальной в данный момент.

На практике уровень актуальности информации часто связывают с характеристиками систем хранения данных. Так, для критически важной информации требуются быстродействующие устройства, но не обязательно большой емкости, а для хранения архивных данных необходимы большие по емкости, но не обязательно быстродействующие средства.

Пусть имеется множество информационных ресурсов  $R = \{R_{ij}\}$ , где в соответствии с приведенной выше классификацией  $i = 1, \dots, n$  – вид ИР;  $j = 1, 2, 3$  – степень актуальности ресурса. Каждый ресурс поступает в информационную систему и проходит в ней весь жизненный цикл.

Представим жизненный цикл единичного ИР  $R_{ij}$  в виде цепи Маркова с дискретным временем (рис. 1). Цепь содержит пять состояний:

- $S_1$  – создание ИР;
- $S_2$  – хранение и обработка ИР критической важности;
- $S_3$  – хранение и обработка ИР с важной информацией;
- $S_4$  – архивирование и хранение ИР с маловажной информацией;
- $S_5$  – удаление ИР.

Вероятности перехода между состояниями ресурса  $i$ -го вида задаются матрицей

$$P_i = \begin{bmatrix} p_{11}^i & p_{12}^i & p_{13}^i & 0 & 0 \\ 0 & p_{22}^i & p_{23}^i & 0 & p_{25}^i \\ 0 & 0 & p_{33}^i & p_{34}^i & p_{35}^i \\ 0 & 0 & 0 & p_{44}^i & p_{45}^i \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

где  $p_{ij}$  – вероятность перехода из состояния  $S_i$  в состояние  $S_j$ ,  $i, j = 1, \dots, 5$ .

Сумма вероятностей в каждой строке матрицы равна единице:

$$\sum_{j=1}^5 p_{ij} = 1, \quad i = 1, \dots, 5.$$

Ресурс характеризуют следующие параметры:

–  $\tau$  – календарное время поступления ИР в информационную систему (например, дни, месяцы),  $0 \leq \tau \leq T$ , где  $T$  – горизонт исследования системы;

–  $T_{si}$  – продолжительность жизни ИР, т. е. количество моментов времени от поступления ресурса в систему до момента его удаления с вероятностью, близкой к единице;

–  $V_{i0}(\tau)$  – объем поступившего ресурса  $i$ -го вида в момент  $\tau$ , Мб;

–  $v_{ij}(t, \tau)$  – объем ресурса  $R_{ij}$ , Мб, в момент  $t \geq \tau$ ;

–  $v_{ij}(t, \tau)$  – величины, образующие вектор  $V_i(t, \tau) = [v_{i1}(t, \tau), \dots, v_{i5}(t, \tau)]$ , при этом

$$\sum_{j=1}^5 v_{ij}(t, \tau) = V_{i0}(\tau), \quad \forall t \geq \tau;$$

–  $x_{ij}(t - \tau)$  – вероятность нахождения ресурса  $R_{ij}$  в  $j$ -й степени актуальности в момент  $t \geq \tau$ . Вероятности  $x_{ij}(t - \tau) = x_{ij}(t, \tau)$  образуют вектор

$X_i(t, \tau) = [x_{i1}(t, \tau), \dots, x_{i5}(t, \tau)]$ , при этом  $\sum_{j=1}^5 x_{ij}(t, \tau) = 1$ ,

а начальное распределение вероятностей  $X_i(\tau, \tau) = [1, 0, 0, 0, 0]$ .

Ресурс, поступивший в систему (в состояние  $S_1$ ), в дальнейшем перераспределяется между состояниями пропорционально вероятностям пребывания системы в данном состоянии. Исходя из вышесказанного, мы получим следующие расчетные формулы.

Динамика изменения состояния ИР определится уравнениями

$$X_i(t+1, \tau) = \begin{cases} O, & t \leq \tau, \\ X_i(t, \tau)P_i, & t \geq \tau, \end{cases} \quad (2)$$

где  $O$  – нулевой вектор,

$$v_{ij}(t, \tau) = \begin{cases} O, & t \leq \tau, \\ V_{i0}(\tau)x_{ij}(t, \tau), & t \geq \tau, \\ i = 1, \dots, n, j = 1, \dots, 5. \end{cases} \quad (3)$$

Общий объем ресурсов  $i$ -го вида, находящихся в системе, задается вектором

$V^{\Sigma}(t) = [V_1^{\Sigma}(t) \quad V_2^{\Sigma}(t) \quad V_3^{\Sigma}(t) \quad V_4^{\Sigma}(t) \quad V_5^{\Sigma}(t)]$ , причем

$$V_i^{\Sigma}(t) = \sum_{\tau=0}^T V_i(t, \tau), \quad t = 0, \dots, T_1, \quad (4)$$

где  $T_1 = T + T_{si}$ , здесь  $T_{si}$  – время жизни ресурса  $i$ -го вида.

Общий объем ресурсов  $j$ -й степени актуальности находится по формуле

$$u_j^{\Sigma}(t) = \sum_{\tau \leq t} \sum_{i=1}^n v_{ij}(t, \tau), \quad i = 2, 3, 4, \quad 0 \leq t \leq T. \quad (5)$$

Расчетный объем ИР  $j$ -й степени актуальности  $U_j^{\Sigma}$  для определения необходимых объемов дискового пространства равен максимальному значению величины  $u_j^{\Sigma}(t)$  на всем исследуемом интервале времени:

$$U_j^{\Sigma} = \max u_j^{\Sigma}(t), \quad j = 2, 3, 4, \quad 0 \leq t \leq T. \quad (6)$$

Расчеты по формулам (2)...(6) для матрицы (1) и исходных данных по объемам поступающих ресурсов удобно производить с помощью разработанной авторами для этой цели программы «Жизненный цикл информационных ресурсов».

Приведем примеры расчетов жизненного цикла ИР на основе предложенной модели и ее программной реализации.

**Модель жизненного цикла единичного ресурса.** Рассмотрим систему, содержащую единственный информационный ресурс.

Предположим, что на вход системы в каждый момент времени поступает ресурс в объеме  $V_{10} = 1$  Мб. Расчет, проведенный по формулам (2)...(5), позволяет оценить динамику накопления ИР в системе (рис. 2).

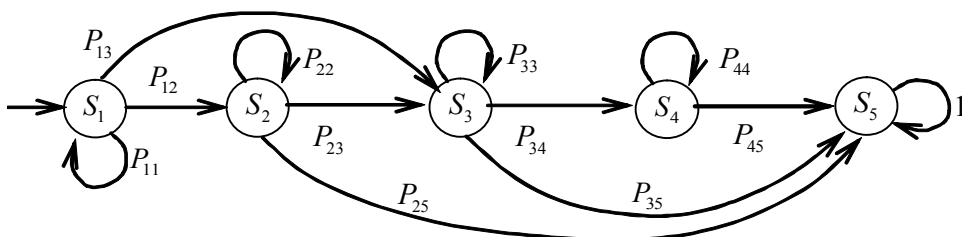


Рис. 1. Вероятностная модель жизненного цикла информационного ресурса

При непрерывном поступлении ИР в систему, которая в начальный момент не содержала ресурсов, процессы поступления и выбытия ИР постепенно уравниваются и информационная система переходит в стационарный режим работы. Установившийся объем ресурсов, относящихся к разным состояниям, различен, за исключением объема удаленных ресурсов (состояния  $S_5$ ), который непрерывно увеличивается (см. рис. 2).

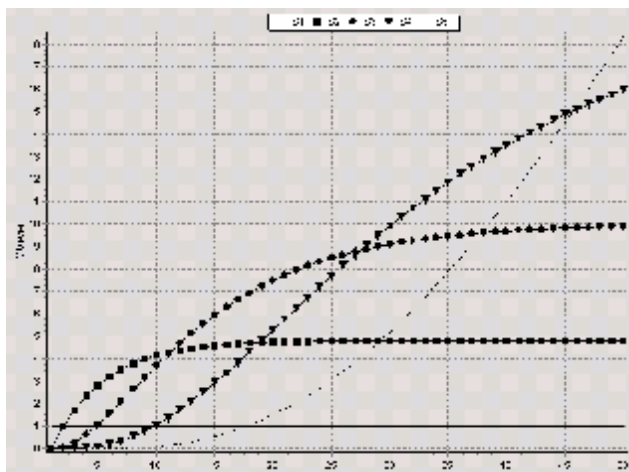


Рис. 2. График динамики накопления ИР при поступлении в каждый момент времени информации объемом 1 Мб

**Модель динамики ИР-организации.** Рассмотрим динамику информационных ресурсов ИТ-организации одного из отделов администрации города. Организация использует информационные ресурсы (в соответствии с приведенной выше классификацией), интенсивность поступления которых зависит от дня недели и квартала. Оценка этих объемов по данным наших наблюдений приведена в таблице.

Основываясь на этих данных, сделаем следующие выводы.

Интенсивность поступления данных и текстовой информации возрастает к концу недели, так как в организацию поступают отчеты из других организаций. Этот же эффект наблюдается в конце каждого квартала, полугодия и года. В конце и начале года интенсивность поступления ИР примерно одинакова, поскольку в этот период принимаются годовые отчеты.

Программное обеспечение обычно обновляется в начале месяца, квартала, года. Обновление производится с помощью сервисных пакетов или путем импорта новых структур.

Интенсивность поступления интерактивных объектов к концу полугодия и года увеличивается, в том числе в связи с необходимостью израсходовать запланированный сетевой трафик.

Что касается ИР типа «Изображение», то здесь не просматривается строгой закономерности их поступления в систему.

Расчет проводится с помощью упомянутой выше программы «Жизненный цикл информационных ресурсов». При работе программы после заполнения матриц  $P_i$  формируются файлы, определяющие сценарий поступления информации по каждому виду ресурсов в течение года в соответствии с приведенной выше таблицей. Один такт работы системы моделирования соответствует одному дню работы организации. Также условно полагаем, что в начале периода моделирования хранилища данных были пусты:

$$V_{i0}(0) = 0, \quad i = 1, \dots, 5.$$

При выполнении расчетов по каждому потоку данных строится график, отображающий вероятность пребывания ресурса в каждом из состояний на каждом такте (рис. 3).

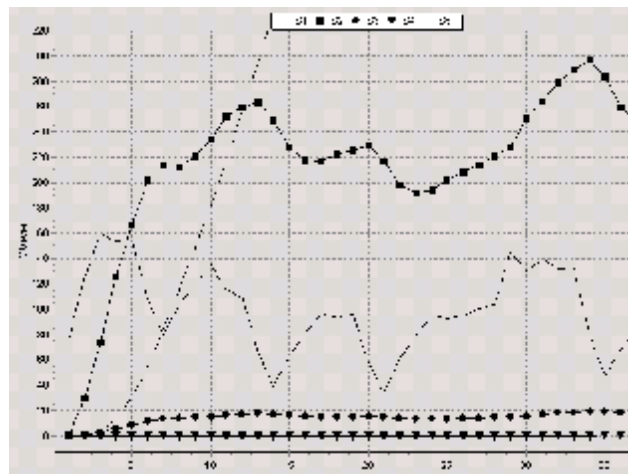


Рис. 3. Динамика ИР «Интерактивные объекты»

С помощью этих графиков можно оценить параметры оборудования, необходимого для хранения информационных ресурсов. Например, на рис. 3 отображено, что ИР «Интерактивные объекты» не передаются на архивное хранение (состояние  $S_4$ ) и практически не хранятся на жестких дисках со средней скоростью доступа (состояние  $S_3$ ). Информационные ресурсы данного типа для хра-

Интенсивность ежедневного поступления данных в течение года, Мб

День (период)	Данные	Изображение	Интерактивный объект	Программные средства	Текст
Понедельник	5	5	40	—	10
Вторник	10	3	43	—	17
Среда	14	5	48	—	21
Четверг	30	2	35	—	35
Пятница	35	2	40	—	35
Конец месяца	75	10	130	—	55
Начало месяца	10	5	40	7	15
Конец квартала	80	15	140	—	80
Начало квартала	10	5	40	7	15
Конец года	75	15	140	—	80
Начало года	110	20	150	7	125

нения используют в основном оборудование с высокой скоростью доступа (состояние  $S_2$ ), после чего сразу удаляются. На графиках также наглядно видна периодичность поступления ИР данного типа.

Резльтирующий график по всем типам информационных ресурсов (рис. 4) показывает, что система выходит на стационарный режим к концу года; в течение года основная нагрузка приходится на жесткие диски архивного хранения; в конце года на них хранится около 1 650 Мб ресурсов всех типов. Менее всего задействованы жесткие диски со средней скоростью доступа: в течение года на них хранится максимум 300 Мб информации.

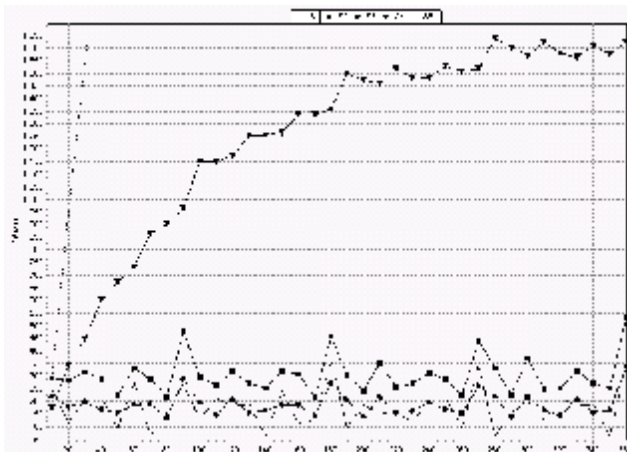


Рис. 4. Динамика суммарного объема ИР всех категорий

После выполнения расчета все исходные и полученные данные можно экспортировать в MS Excel для дальнейшей обработки и анализа. Программа позволяет обрабатывать от одного до пяти типов ресурсов на глубину до 1 000 000 тактов. В ней также реализован контроль ввода и предусмотрена возможность сохранения исходных данных.

Таким образом, предложенная в данной статье вероятностная модель жизненного цикла информационных ресурсов позволяет получать качественно адекватную картину динамики как для отдельного ресурса, так и для совокупности ресурсов, используемых ИТ-организацией.

С помощью программы «Жизненный цикл информационных ресурсов», реализующей эту методику, мож-

но проанализировать загруженность дискового оборудования несколькими типами ресурсов за любой срок и, исходя из полученных данных, подобрать оптимальное по цене, объему и скорости доступа оборудование.

Кроме того, можно решать задачу по поиску оптимальной стратегии по управлению мощностями [1], рассматривая сроки замены и характеристики оборудования.

Определенную сложность при использовании предлагаемой методики представляет получение оценок вероятностей перехода между состояниями  $p_{ij}$ . Для этой цели, помимо рассмотренного выше метода экспертного оценивания, можно использовать методику оценки параметров цепей Маркова по экспериментальным данным [8].

### Библиографический список

1. Ниессен, Я. Введение в ИТ Сервис-менеджмент [Электронный ресурс] / Я. Ниессен. Электрон. дан. Режим доступа: <http://allitil.ru/book/42-vvedenie-v-itsm-glava-1.html>. Загл. с экрана.
2. Антопольский, А. Б. Информационные ресурсы России : науч.-метод. пособие / А. Б. Антопольский. М. : Либерия, 2004.
3. Избачков, Ю. И. Информационные системы : учебник для вузов / Ю. С. Избачков, В. Н. Петров. СПб. : Питер, 2005.
4. Hierarchical Storage Management [Electronic resource]. Electronic data. Access mode: [http://www.dscon.ru/education/legato\\_hsm\\_2.htm](http://www.dscon.ru/education/legato_hsm_2.htm). Title from display.
5. Dublin Core Metadata Element Set. Version 1.1 : Reference description [Electronic resource]. Electronic data. 1999. Access mode: <http://dublincore.org/documents/dces/>. Title from display.
6. American Productivity&Quality Center [Electronic resource]. Electronic data. Access mode: <http://www.apqc.org/portal/apqc/site.html>. Title from display.
7. Жукова, С. А. Моделирование и оптимизация структуры информационного ресурса : автореф. дис. ... канд. техн. наук / С. А. Жукова. Ижевск, 2008.
8. Доррер, Г. А. Технология моделирования и разработки учебных электронных изданий / Г. А. Доррер, Г. М. Рудакова. Новосибирск : Изд-во Сиб. отд-ния Рос. акад. наук, 2006.

G. A. Dorrer, A. A. Popov, K. V. Sysenko

## LIVING CYCLE INVESTIGATION OF THE ELECTRONIC INFORMATION RESOURCES

*A probability model of information resources (IR) living cycle based on the Markov chain formalism is suggested. Here are given the following initial suppositions:*

- IR are classified according to the Dublin Core Metadata Element Set (DCMI);
- five typical business processes on IR managing according to American Productivity & Quality Center classification are take out;
- IR are classified as critically important, important and non-important.

*With the help of the developed program system the proposed model can estimate the living cycle of all the components of the single IR, their dynamics in organization of newly entered IR. It gives the opportunity to estimate the necessary disc space capacity in organization. The sample of IR dynamics modeling in some department of the city administration is given.*

*Keywords: information recourses, living cycle, simulation.*