

Информационные ресурсы и технологии

ШИШАЕВ Максим Геннадьевич – кандидат технических наук, заведующий лабораторией Института информатики и математического моделирования технологических процессов Кольского научного центра РАН, заместитель декана Кольского филиала Петрозаводского государственного университета.

ДАТЬЕВ Игорь Олегович – младший научный сотрудник Института информатики и математического моделирования технологических процессов Кольского научного центра РАН

ШЕМЯКИН Алексей Сергеевич – аспирант Института информатики и математического моделирования технологических процессов Кольского научного центра РАН

МЕТОДИКА ОЦЕНКИ УРОВНЯ ИНФОРМАЦИОННЫХ ПОТРЕБНОСТЕЙ РЕГИОНАЛЬНЫХ ПОЛЬЗОВАТЕЛЕЙ*

Оценка объемов информационных потребностей региональных пользователей является актуальной задачей. Прежде всего, знания об объемах потребностей пользователей необходимы для эффективного развития регионального рынка информационных услуг. В этом контексте особенно важным является получение перспективных оценок, способствующих формированию наилучших стратегий развития компаний, предоставляющих информационные услуги. Кроме того, в современных условиях такие оценки необходимы и для эффективного решения задач прогнозирования и планирования стратегий информационного развития регионов в целом, анализа существующего состояния информационных и телекоммуникационных услуг, разработки и внедрения новых услуг и информационных технологий [1, 2].

Под информационными потребностями здесь понимается совокупный объем информационно-вычислительных работ, возникающий в связи с социальной и экономической деятельностью пользователя. Информационная потребность имеет несколько аспектов: потребность в хранении информации, ее обработке и потребность в информационном обмене (сетевая информационная потребность). В данной работе будет рассматриваться, в основном, последний аспект.

Количественная оценка объема пользовательских информационных потребностей является, по сути, задачей анализа рынка информационных услуг, которую можно интерпретировать как оценку объема рынка или оценку спроса [3]. В настоящее время сложились три основных подхода к решению данной задачи:

1. непосредственное изучение параметров соответствующего сектора рынка (объемы и структура продаж товаров и услуг, объемы потребленного телекоммуникационного трафика и т.п.);
2. анализ объема и структуры рынка с помощью опросов (анкетирование целевых групп);
3. оценка объемов спроса на базе анализа вторичной информации и экспертных оценок.

Первые два подхода в настоящее время достаточно широко используются на практике (см., например, [4]), однако их применимость в ряде случаев ограничена. Главной проблемой первого является необходимость сбора больших объемов специфической информации, получение которой, во-первых, трудоемко, а во-вторых, часто бывает ограничено соображениями коммерческой тайны. Второй подход также весьма трудоемок и накладывает условие репрезентативности на опрашиваемую аудиторию. Важной проблемой является также и скорость получения оценки – если процедура анализа рыночных показателей или проведения опроса затягивается, то появляется риск того, что полученная информация не будет соответствовать реальной ситуации, сложившейся на рынке.

Кроме оценки текущего состояния рынка важной задачей является также получение количественных прогнозов его развития. Прогнозирование рынков, так или иначе, опирается на апостериорный анализ статистических данных об экономических или иных параметрах рынка с последующей экстраполяцией результатов на период перспективной оценки. Для реализации такого подхода в настоящее время существует достаточно большое количество методик [3, 5]. Главной проблемой при этом является необходимость оперирования большими объемами статистических данных за длительный период времени (рекомендуемый в методиках экономического анализа период – не менее 10 лет).

Таким образом, в условиях ограниченной доступности достоверных статистических данных о рынке информационных услуг, а также высоких требований к оперативности получения результатов, наиболее приемлемым способом определения потребности пользователей в информационных ресурсах сети оказывается построение эвристической детерминированной вычислительной модели объемов информационных потребностей, основанной на легкодоступных (в смысле скорости и стоимости доступа) данных. Подобная модель расчета, разумеется, не будет обладать высокой точностью, однако будет существенно превосходить другие методы получения оценок по скорости и стоимости получения результата.

* Работы выполнены по программе ОИТВС РАН "Фундаментальные основы информационных технологий и систем" (проект № 2.6) и поддержаны грантом РФФИ (проект № 05-07-90050)

Предложенная в данной работе методика позволяет оценить объем регионального рынка предоставления Интернет-услуг в целом, рассчитать значения отдельных показателей регионального рынка информационных услуг при различных значениях исходных параметров технического и экономического свойства. Модель была использована для оценки перспективных сетевых информационных потребностей пользователей Мурманской области. Правдоподобность полученных с ее помощью оценок была подтверждена проверкой количественных показателей с помощью альтернативных методов расчета.

Исходные данные

В качестве исходных данных для оценки сетевых информационных потребностей пользователей региона использовались социально-экономические показатели городов и районов Мурманской области, предоставленные Территориальным Органом Федеральной Службы Государственной Статистики по Мурманской Области за 2004 год [6]. В расчетах были использованы данные о численности населения, численности и структуре занятого населения, распределении занятого населения по отраслям экономики. Кроме того, в качестве исходных данных в расчетной модели использовались дополнительные параметры, определяющие связь различных социально-экономических показателей с объемами и структурой информационных потребностей. Для получения количественных значений параметров использовались экспертные оценки и информация из открытых источников интернета. К использованным дополнительным параметрам относятся:

1. Количественная оценка месячной удельной базовой информационной потребности одного пользователя.
2. Отраслевые коэффициенты информационной потребности («информациоёмкость» отрасли), определяющие уровень информационной ёмкости процессов, связанных с различными отраслями региональной экономики.
3. Уровень «интернетизации» населения по районам и городам – соотношение населения, пользующегося услугами интернета, ко всему населению.
4. Соотношение корпоративных и частных пользователей интернета. Под корпоративными подразумеваются все пользователи, которые используют интернет в связи со своей служебной деятельностью, частные пользователи используют сетевые ресурсы в личных целях.

Методика расчета

Описанная в данной работе методика расчета сетевых информационных потребностей опирается на идеи, изложенные в работе [7]. Основным ее постулатом является предположение о том, что количественные характеристики информационных процессов, ассоциированных с некоторой социально-экономической системой (СЭС), зависят от внутренней структуры и значений параметров этой системы. При этом величина оценки объемов информационных процессов в региональной СЭС в целом определяется как сумма аналогичных оценок ее компонентов. Каждая из последних, в свою очередь, зависит от типа компонента и его объемной характеристики: для каждого типа компонента определяется коэффициент его удельной информациоёмкости, и искомая оценка объема информационных процессов вычисляется как произведение коэффициента информациоёмкости на значение объемной характеристики данного конкретного элемента системы. Например, если в качестве элементов социально-экономической системы рассматривать отдельные отрасли экономики, то для проведения расчетов необходимо располагать коэффициентом информациоёмкости для каждой из них, а в качестве их объемной характеристики можно использовать количество занятых в данной отрасли или приходящуюся на нее долю ВВП региона.

В качестве компонентов, используемых в расчетной модели, можно использовать территориальные (районы, области) и функциональные (отрасли экономики) элементы социально-экономической системы различного уровня детализации. При делении по функциональному признаку на самом общем уровне это могут быть отрасли экономики. В этом случае методика, разумеется, применима лишь для систем большого масштаба и способна дать лишь весьма общие оценки, однако для крупных элементов, как правило, проще оценить их удельную информациоёмкость. Более точные количественные оценки информационных процессов могут быть получены при оперировании элементами СЭС масштаба предприятия, однако оценка информациоёмкости отдельно взятого промышленного объекта является сама по себе весьма трудоемкой и дорогостоящей задачей.

Расчет информационных потребностей элементов социально-экономической системы может вестись двумя способами.

Распределительный метод может быть использован, когда известно значение информационной потребности пользователей для крупного региона (региона $m-1$ -го уровня) и требуется распределить это значение между его подрегионами (регионами m -го уровня). Объем информационной потребности отдельно взятого m -го региона определяется по формуле:

$$V_m = \frac{V_{m-1} \cdot \theta_m(X_1, X_2, \dots, X_\alpha)}{\theta_{m-1}(X_1, X_2, \dots, X_\alpha)} \quad (1)$$

Здесь V_{m-1} – объем информационных потребностей вышележащего региона, X_i – социально-экономические параметры, коррелирующие с определяемым показателем информационных потребностей, а θ_m, θ_{m-1} – функции множественной регрессии.

Для практических целей функции множественной регрессии могут быть заменены их линейными аппроксимациями. Кроме того, для быстрого получения приближенных оценок вместо множества социально-экономических параметров в расчетной формуле можно использовать некоторую агрегированную объемную характеристику рассматриваемых регионов. В этом случае для некоторой крупной СЭС должны быть известны оценка суммарной информационной потребности ассоциированных с ней пользователей и коэффициенты удельной информативности ее элементов. Тогда для территориальных элементов рассматриваемой системы объем информационных потребностей пользователей рассчитывается пропорционально информационным потребностям входящих в них функциональных элементов. Последние, в свою очередь, определяются исходя из коэффициентов информативности элементов СЭС. Для случая информационных потребностей функциональных элементов социально-экономической системы (отраслей) расчетная формула будет иметь вид:

$$P_i = k_i \frac{V_i}{V} P \quad (2),$$

где P_i – искомый объем информационных потребностей пользователей, ассоциированных с i -й отраслью рассматриваемого региона,

k_i – удельный коэффициент информативности i -й отрасли;

V_i – объемная характеристика i -й отрасли рассматриваемого территориального элемента социально-экономической системы;

V – суммарная объемная характеристика территориального элемента;

P – оценка суммарной информационной потребности региональных пользователей.

Прямой метод расчета позволяет обойтись без оценки суммарной информационной потребности – вместо этого в качестве базового показателя, на основе которого рассчитываются информационные потребности функциональных и территориальных элементов социально-экономической системы, используется усредненная базовая информационная потребность одного пользователя. Информационная потребность региона вычисляется как сумма информационных потребностей пользователей, ассоциированных с различными отраслями экономики, а информационная потребность региона более высокого уровня (области) – как сумма соответствующих потребностей его подрегионов (районов и городов).

Сетевая информационная потребность отрасли экономики в одном регионе вычисляется по следующей формуле:

$$T_{ij} = U_{ij} C_j B \quad (3),$$

где

T_{ij} – информационная потребность пользователей интернета i -ой отрасли экономики в j -ом регионе (городе или районе),

U_{ij} – количество пользователей интернета, занятых в i -ой отрасли экономики в j -ом регионе,

C_i – коэффициент информативности i -ой отрасли,

B – базовая информационная потребность.

Количество пользователей интернета, занятых в i -ой отрасли в j -ом регионе вычисляется по следующей формуле:

$$U_{ij} = P_j Q_j R_j S_{ij} \quad (4),$$

где

U_{ij} – количество пользователей интернета, занятых в i -ой отрасли экономики в j -ом регионе,

P_j – население j -го региона,

Q_j – коэффициент «интернетизации» для j -го региона, равный отношению количества пользователей интернета ко всему населению региона.

R_j – доля корпоративных пользователей интернета среди всех пользователей интернета.

S_{ij} – доля Интернет-пользователей, работающих в j -ом регионе в i -ой отрасли.

Следует отметить, что если рассматривать используемые в данных расчетных формулах коэффициенты, соответствующие параметрам социально-экономического развития региона, как функции времени, то расчетная модель становится динамической. Это позволяет использовать ее для получения не статичных, но перспективных оценок информационных потребностей. Однако аналитическое выражение подобных временных функций представляется, по меньшей мере, затруднительным, если вообще возможным. В то же время, для получения значений коэффициентов, отображенных на различные точки временной оси, могут быть с успехом использованы методы и технологии имитационного моделирования. Такое применение приведенной расчетной методики (с небольшими модификациями) было реализовано в

Институте информатики КНЦ РАН в рамках работ по созданию системно-динамических моделей региональной информатизации [8, 9].

Пример применения методики

Для определения информационной потребности пользователей интернета Мурманской области был использован прямой метод расчета. В расчетах использовались следующие значения исходных параметров:

1. Базовая информационная потребность – 100 Мбайт/мес. Это значение является результатом экспертной оценки, полученной на основе опроса частных пользователей интернета, а также анализа усредненных значений месячного трафика пользователей нескольких корпоративных сегментов интернета Мурманской области.
2. Количественные оценки коэффициентов информативности различных отраслей основываются на данных, приведенных в работе [7]. Несмотря на то, что указанный источник несколько устарел, было сделано предположение, что соотношения значений коэффициентов, соответствующих отдельным отраслям, не претерпели существенных изменений. Были использованы следующие числовые значения коэффициентов:
 - a. Промышленность – 2,5
 - b. Сельское хозяйство – 0,1
 - c. Торговля – 0,2
 - d. Транспорт и связь – 2
 - e. Строительство – 0,2
 - f. ЖКХ – 0,5
 - g. здравоохранение, социальное обеспечение – 0,7
 - h. Образование, культура, наука – 1,5
 - i. Домашние пользователи – 1
3. Коэффициент интернетизации населения Q (см. формулу (4)) был вычислен на базе данных по статистике российской части интернета за 2005 год [10]. Значение коэффициента - 3,67%.
4. В качестве соотношения корпоративных и домашних пользователей (коэффициент R в формуле (4)) использовалась экспертная оценка, равная 50%.
5. Распределение работающего населения по отраслям экономики Мурманской области вычислено на основе данных, предоставленных Территориальным Органом Федеральной Службы Государственной Статистики по Мурманской области [6].

Результаты вычислений приведены в таблице 1.

Таблица 1

Значения информационной потребности городов и районов Мурманской области по отраслям экономики

Информационная потребность пользователей, Мбайт\мес	Промышленность	Транспорт и связь	Здравоохранение, соц. обесп.	Образование, культура, наука	Домашние пользователи	Прочее	Итого по городам
Мурманск	271 434	233 851	43 847	128 857	596 559	41 550	1 316 098
Апатиты	35 718	17 423	14 148	39 899	116 156	9 507	232 851
Кандалакша	50 002	52 905	8 656	27 904	107 531	5 823	252 822
Кировск	98 420	1 747	3 581	9 826	62 365	2 969	178 909
Мончегорск	109 872	6 504	8 484	20 545	98 540	5 390	249 334
Оленегорск	68 313	963	4 592	15 348	60 188	4 015	153 419
Полярные Зори	44 008	549	2 474	6 588	34 315	1 994	89 928
Ковдорский р-он	61 278	780	2 152	9 354	43 306	2 018	118 888
Кольский р-он	42 915	18 864	5 018	23 485	94 319	7 267	191 868
Ловозерский р-он	21 714	1 418	1 383	6 647	25 323	1 662	58 148
Печенгский р-он	106 525	4 706	4 471	16 010	84 043	4 715	220 470
Терский р-он	3 211	1 927	908	3 449	12 845	973	23 313
Итого по отраслям	913 411	341 637	99 715	307 914	1 335 488	87 884	3 086 048

Заключение

С целью верификации полученной модели было произведено сравнение результирующих данных построенной модели и реальных статистических данных, предоставленных одним из провайдеров Мурманской области – компанией «Релант».

Согласно предоставленным данным, объем трафика, ассоциированного с пользователями, составляет примерно 125 гигабайт в месяц. По предварительным оценкам, доля пользователей «Релант» составляет приблизительно 5% от всех пользователей интернета Мурманской области. При таких предположениях, информационная потребность пользователей интернета составляет 2,5 терабайт в месяц. Это значит, что разница между реальными данными и данными, полученными с помощью вычислительной модели, не превышает 20%.

Таким образом, можно выделить, по крайней мере, два основных достоинства предлагаемой методики:

1. использование в расчетах в основном вторичной информации, получение доступа к которой не представляет особой сложности;
2. возможность интеграции расчетной методики с системами имитационного моделирования социально-экономических систем с целью получения перспективных оценок информационных потребностей пользователей (прогнозирования).

Литература

1. Олейник А.Г., Шишаев М.Г. Пути повышения эффективности использования региональных информационных ресурсов// Информационные ресурсы России. – 2004. - №1.- С. 2-5.
2. Олейник А., Шишаев М. Нетехнологические аспекты региональной информатизации// Информационные технологии в региональном развитии: концептуальные аспекты и модели. – Апатиты, 2002. – С.25-30.
3. Баканов М.И., Мельник М.В., Шеремет А.Д. Теория экономического анализа. – М.: Финансы и статистика, 2006. - 536 с.: ил.
4. Горбунова З.В., Карлина Ю.О. Модель расчета доходов от услуг широкополосного доступа // Информационные ресурсы России. - 2005. - №5. – С.36-37.
5. Аналитические технологии для прогнозирования и анализа данных. Учебные материалы компании «НейроПроект». Документ: <http://www.neuroproject.ru/tutorial.php>. Дата доступа 03.10.2006.
6. Города и районы Мурманской области// Федеральная служба государственной статистики, территориальный орган Федеральной службы государственной статистики по Мурманской области. - Мурманск, 2005. -150 с.
7. Основы построения больших информационно-вычислительных сетей. Под общей ред. Д.Г. Жимерина и В.И.Максименко. - М.: Статистика, 1976, - 296 с., ил.
8. Шишаев М.Г., Фомин Д.В. Опыт построения СД-модели Интернет-трафика региона // Информационные технологии в региональном развитии. – Апатиты, 2004. – Вып. IV. – С. 31-38.
9. Шишаев М.Г., Олейник А.Г. СД-модели для оперативного анализа организации распределенных вычислений // Информационные технологии в региональном развитии. – Апатиты, 2004. – Вып. IV. – С. 39-42.
10. Статистика Рунета за Апрель 2005. Документ: <http://itpromotion.ru/articles/promo/article4.html#five>. Дата доступа 03.10.2006.

11.

Шишаев Максим Геннадьевич - кандидат технических наук, заведующий лабораторией региональных информационных систем Института информатики и математического моделирования технологических процессов Кольского научного центра РАН, заместитель декана факультета информатики и прикладной математики Кольского филиала Петрозаводского государственного университета.



Датьев Игорь Олегович – младший научный сотрудник Института информатики и математического моделирования технологических процессов Кольского научного центра РАН



аспирант Института
моделирования научного
Кольского центра

Шемякин Алексей Сергеевич –
информатики и математического
технологических процессов
центра РАН

