

Program of automated keyword generation based on a collection of web documents

Alexander Sergeevich Zhuravel, Developer, Wikimart Technologies Ltd.

The paper is about two approaches to the problem of keyword generation based on a collection of web documents. A new approach is developed on the basis of the analysis.

Keywords: Context advertising, keyword generation, TF-IDF, analysis of a web document.

УДК 378.1

ГЕОСТАТИСТИЧЕСКИЙ АНАЛИЗ В ОБРАЗОВАНИИ

Ольга Викторовна Зайцева, канд. техн. наук, зав. отделом статистики
Центра мониторинга и статистики образования,
E-mail: cvdisser@list.ru,
Федеральный институт развития образования,
<http://www.firo.ru>

Статья описывает новое научное направление – геостатистику. Показана связь геостатистики с научной картиной мира. Показано, что как прикладное направление геостатистика создает механизм исследования и управления системой образования. Исследован аспект применения геостатистики в образовании. Показано, что геостатистика создает информационные поля, которые связывают воедино разрозненные пространственные объекты одного качества, особенности геостатистики. Показано, что геостатистика в образовании отражает тенденцию информатизации образования.

Ключевые слова: образование, геоинформатика, статистика, геостатистика, моделирование, визуализация, геоданные, принципы геостатистики, управление образованием.

Введение

Научная картина мира – одно из основополагающих понятий современной науки, которое включает обобщение и синтез различных научных теорий [1]. Научная картина мира включает общенаучную картину мира и картины мира отдельных наук. Картины мира отдельных наук включают в себя определённые способы понимания и трактовки предметов, явлений и процессов объективного мира, существующие в каждой отдельной науке. Научная картина мира выступает как специфическая форма систематизации научного знания, задающая видение предметного мира науки соответственно определенному этапу её функционирования и развития [2].



О.В. Зайцева

Наряду с понятием «научная картина мира» и «картины мира отдельных наук» применяют термин «картина мира». В философско-методологической литературе термин «картина мира» применяется как для обозначения мировоззрения, так и в более узком смысле, когда речь заходит о таких представлениях об окружающей действительности, которые являются особым типом теоретического или практического знания. Картина мира в рамках какой-либо отрасли [3] включает понятие информационного пространства этой отрасли. Информационное пространство является отражением информационной сферы Земли [4]. Для построения такого пространства, в частности в сфере образования, необходимо использовать статистические методы. Таким образом, для построения описания [5] отраслевой картины мира, необходимо применять статистические методы

и геостатистические методы. При этом систему образования можно рассматривать как сложную организационно-техническую систему [6], что позволяет использовать известные модели и методы описания и применения таких систем. Таким образом, геостатистика связана с «научной картиной мира» через «картину мира». Как прикладное направление геостатистика создает механизм исследования и управления системой образования, что определяет актуальность исследования в данной области.

Эволюция геостатистики. Для привязки информационного пространства к реальному пространству необходимо применять геостатистику, как науку о статистическом описании реальных пространственных процессов объектов и явлений. При этом возможны два подхода. Первый основан на построении информационных моделей явлений, процессов и систем [7], с последующей привязкой к статистической информации и статистической динамике. Он применяет обычную статистику. Вторым подходом основан на использовании научного направления называемого геостатистика [8], которое включает синтез параметров описывающих предметные информационные поля со статистической информацией об этих полях и построение таких синтезированных моделей.

Геостатистика является специфическим научным направлением. Длительное время она не признавалась статистиками как наука, и поэтому ошибочно ее связывать в первую очередь со статистикой. С появлением и развитием геоинформатики геостатистика получила новое развитие [9]. В определенной степени геостатистику можно сравнивать с операционным исчислением. Первоначально были созданы математические рабочие формулы, и через какое-то время была создана теория. Первые работы в этой области связывают с работами в области геологии [10] горнодобывающей промышленности, а также гидрологии [8]. Следует также отметить, что приставка «гео» (геоданные) длительное время связывалась с геологией, геодезией и в меньшей степени с географией. «Геоданные» трактовались как «геологические данные» или «геодезические данные», то есть специальные данные в определенной предметной области.

С появлением геоинформатики «геоданные» трактуют как «геоинформационные данные» или интегрированные данные, применяемые в разных областях: геодезии, геологии, картографии, инженерных изысканиях [11], дистанционном зондировании и др.

В настоящее время геостатистика имеет два значения [12]:

- 1) отображение статистической информации методами геоинформатики, включая сферу образования;
- 2) направление построения геоинформационных моделей и геоинформационного моделирования (геомоделирования) на основе специфических вероятностных методов, истоки которых восходят к Герхарду Криге и французскому математику Матерону [13].

Основная идея геостатистического метода изложена в магистерской диссертации Герхарда Криге, но развита позже Матероном. Она выглядит несколько вульгарно для статистиков и математиков. Берется симметричная вариограмма, отбрасывается ее половина, и ищется функция, обнуляющая эту половину на эмпирически измеренных точках. Это направление геостатистики позволяет создавать непрерывные информационные поля по совокупности измеренных точечных значений [9, 13].

Геостатистика в образовании. Геостатистика в образовании позволяет решать ряд задач. Первая задача (технологическая) – это описание системы образования как сложной организационно-технической системы [6] для того, чтобы можно было бы эффективно ею управлять. Вторая задача (научная) – это описание системы образования как «картины мира» [14], для изучения и научного анализа этой системы. Третья задача (ситуационная) – отражение ситуации в сфере образования и описание существующих в ней социальных и экономических отношений [15]. Четвертая задача (управленческая) – анализ и совершенствование механизмов и технологий управления в сфере образования

[16, 17]. Пятая задача (ресурсная) – это исследование информационных образовательных ресурсов с целью их накопления и использования [18].

Геостатистика в управлении образованием представляет собой технологию поддержки принятия решений. Она опирается на первое, третье, четвертое и пятое направление геостатистики. В этом аспекте ее можно рассматривать как инновационную геоинформационную технологию. Современное развитие образования широко использует инновационные технологии и образовательные инновации для улучшения качества образовательных услуг.

Принципиальным моментом применения геостатистики в образовании является то, что она использует информационное пространство для формирования информационного поля [19]. Геостатистика в образовании ложится в русло информатизации образования. Информатизация образования направлена на создание информационных образовательных ресурсов, повышение качества образования, инновационное развитие образования и увеличение интеллектуального капитала [20] в сфере образования.

Геостатистика в управлении образованием может быть рассмотрена как система, включающая следующие основные факторы: концептуальные, организационные, информационные, управленческие, человеческие. Рассматривать отдельно факторы геостатистики по нашему мнению нецелесообразно, из-за высокой степени интеграции современной образовательной информационной среды.

Геостатистика использует пространственные данные или геоданные [21]. Геоданными называют данные о процессах и явлениях на земной поверхности, которые включают три классифицированные и интегрированные в единую систему группы данных: «место», «время», «тема».

Подчеркнем, что геоданные, как обобщение данных, включают не только данные области наук о Земле, но и других областей. К этим дополнительным областям относят: транспорт, экономику, экологию, управление, образование, анализ, искусственный интеллект и т. д. Другими словами, объем понятия термина «геоданные» в современной трактовке вышел за рамки наук о Земле. Этим можно отразить лингвистическую особенность геоданных [21].

Системная особенность геоданных состоит в том, что после их формирования они представляют собой систему, связывающую и согласовывающую данные разных типов и структур в единый системный комплекс [21]. Информационная особенность геоданных состоит в том, что они представляют собой новый информационный ресурс, который позволяет решать задачи разных предметных областей [21].

Локализационное значение геоданных связано с локализацией их в реальном пространстве. Существуют различные зависимости между расположением и значением данных. Поэтому понятие «локализация» имеет, по крайней мере, два значения [22]: одно – это положение точки в пространстве и второе – пространственный объект (плоский или объемный), привязанный к точке пространства. В геостатистике применяется второе значение локализации. При этом значение локализованных данных, связанных с областью может модифицировано как среднее значение наблюдаемой величины, среднее значение на площадь области, значение на объем области. В последнем случае площадь или объем часто называют «поддержкой» данных [8, 10]. Это тесно связано с идеей поддержки измерений или поддержки принятия решений.

Если связать площадь с точкой пространства, то она может служить выражением различных характеристик объектов, которые расположены на данной площади. Такие объекты могут принадлежать сфере образования. Этот прием применяют в классической статистике, когда локальной точке – центру субъекта Федерации РФ или центру более мелкого объекта присваивают статистический показатель. Например, среднее количество школ или общее количество школ в регионе, количество

преподавателей на школу в регионе, количество учеников в регионе, средний бал по предмету и прочее.

Эти статистические показатели с помощью геоинформационных методов переносят на визуальную модель (электронную карту). Таким образом, визуальная геостатистическая модель представляет собой геоинформационную визуализацию. Геостатистическое моделирование является частным случаем геоинформационного моделирования.

Геостатистика имеет особенности методологии и принципов изучения явлений. Как статистика она изучает массовые количественные отношения, распределенные в пространстве в неразрывной связи с качественными особенностями процессов и явлений, но как раздел геоинформатики она исследует пространственные взаимосвязи и пространственные отношения [15] между объектами и явлениями.

При этом она анализирует дифференциацию и интеграцию как пространственных, так и временных процессов и явлений. Таким образом, статистические характеристики тесно связываются с пространственными отношениями, и в результате геостатистика показывает совокупность явлений и их развитие в пространственном и графическом представлении.

Геостатистика позволяет изучать временные факторы процессов, протекающих в сфере образования. При изучении временных процессов подход в геостатистике состоит в том, что функция, включающая временные зависимости может быть представлена суммой следующих компонентов: структурный компонент (тренд), циклический компонент, конъюнктурный компонент и случайный шум. В этом аспекте это понятие такой функции близко к понятию *временного ряда*.

Для исследования явлений и процессов методами геостатистики следует, прежде всего, собрать необходимые сведения – геостатистические данные. Под геостатистическими данными понимают совокупность количественных характеристик массовых явлений и процессов, связанных пространственными отношениями. Геостатистические данные можно рассматривать как разновидность геоданных, характеризующих статистические процессы. Поэтому в геоинформатике геостатистические данные включают в более общее понятие – геоданные.

Геостатистическое наблюдение (измерение) – это организованная процедура, которая заключается в фиксации и регистрации определенных характеристик объекта исследования, входящего или в некое множество.

Цель геостатистического наблюдения – получение достоверной информации для выявления закономерностей развития массовых явлений и процессов, имеющих пространственную распределенность и находящихся друг с другом в неких пространственных отношениях.

Сбор геоданных осуществляется по двум каналам пространственная информация и статистическая информация. В одном технологическом варианте собирается только статистическая информация, которая затем классифицируется и геокодируется. Это пример сбора обычной статистики. В другом, собирается пространственная информация, сразу же содержащая статистические характеристики. Совокупность показателей, отображающих объект, явление или процесс образует систему показателей. Система показателей – модель, созданная с целью описания и изучения взаимозависимости между массовыми явлениями или процессами.

Целью геостатистической обработки является обобщение информации и получение оценок параметров, позволяющих анализировать и интерпретировать данные, получаемые при исследовании сферы образования.

Выводы. Геостатистика позволяет получать новое знание, которое называют пространственным знанием [23]. Пространственное знание позволяет решать новые задачи в области геоинформатики, в области искусственного интеллекта и в области наук о Земле. Пространственное знание применяется при создании, организации и

управлении отраслями и территориальными комплексами. Геоestatистика в управлении образованием представляет собой инновационную технологию, позволяющую решать задачи управления образованием, сбора информации, построения уникальных информационных полей, создания информационных образовательных ресурсов и информационных ресурсов управления. Как технология управления она относится к технологиям поддержки принятия решений. Геоestatистика позволяет получать новое знание, и это знание классифицируется как пространственное знание.

Литература

1. *Леонтьев А.Н.* Образ мира // Избранные психологические произведения. М.: Педагогика, 1983. С. 251–261.
2. *Степин В.С., Кузнецова Л.Ф.* Научная картина мира в культуре техногенной цивилизации. М.: ИФ РАН, 1994. 274 с.
3. *Цветков В.Я.* Картина мира как образовательная парадигма // *European Social Science Journal* = Европейский журнал социальных наук. 2013. № 10–1 (37). С. 28–34.
4. *Соловьёв И.В.* Проблемы исследования информационной сферы Земли // *Перспективы науки и образования*. 2014. № 5. С. 25–30.
5. *Цветков В.Я.* Информационное описание картины мира // *Перспективы науки и образования*. 2014. № 5. С. 9–13.
6. *Цветков В.Я.* Информационные модели объектов, процессов и ситуаций // *Дистанционное и виртуальное обучение*. 2014. № 5. С. 4–11.
7. *Кужелев П.Д.* О применении геоestatистики в науках о Земле // *Науки о земле*. № 4. 2012. С. 77–81.
8. *Майоров А.А., Матерухин А.В.* Геоинформационный подход к задаче разработки инструментальных средств массовой оценки недвижимости // *Геодезия и аэрофотосъемка*. 2011. № 4. С. 92–97.
9. *Agterberg F.P.*, *Geomathematics, Mathematical Background and Geo-Science Applications*, Elsevier Scientific Publishing Company. Amsterdam, 1974.
10. *Цветков В.Я.* Геоestatистика // *Геодезия и аэрофотосъемка*. 2007. № 3. С. 174–184.
11. *Matheron G.* The intrinsic random functions, and their applications, *Adv. Appl. Prob.*, 5. P. 439–468, 1973.
12. *Tsvetkov V.Ya.* Worldview Model as the Result of Education // *World Applied Sciences Journal*. 2014. 31 (2). P. 211–215.
13. *Цветков В.Я.* О пространственных и экономических отношениях // *Международный журнал экспериментального образования*. 2013. № 3. С. 115–117.
14. *Ожерельева Т.А.* Применение информационных технологий для управления образовательными процессами // *Управление образованием: теория и практика*. 2013. № 4. С. 133–137.
15. *Зайцева О.В.* Геоestatистика в управлении образованием // *Управление образованием: теория и практика*. 2013. № 3. С. 47–51.
16. *Tsvetkov V.Ya.* Information field. *Life Science Journal*. 2014. 11(5). P. 551–554.
17. *Зайцева О.В.* Информатизация образования и интеллектуальный капитал // *Дистанционное и виртуальное обучение*. 2012. № 12. С. 105–109.
18. *Савиных В.П., Цветков В.Я.* Геоданные как системный информационный ресурс // *Вестник Российской Академии Наук*. 2014. Т. 84. № 9. С. 826–829.
19. *Цветков В.Я.* Геоинформационные системы как системы пространственно-локализованных данных. М.: ГосНИИ ИТТ «Информика», 1999. 113 с. [Электронное издание]. Номер гос. регистрации 0329900095.
20. *Antony Galton.* Spatial and temporal knowledge representation // *Earth Science Informatics*. 2009. September. Vol. 2. Issue 3. P. 169–187.

Geostatistical analysis in education

Olga Viktorovna Zaitseva, Candidate of Technical Sciences. Head of the Department of Statistics and Monitoring Center for Education Statistics, Federal Institute for Educational Development, Moscow, Russia.

The article describes a new scientific direction – geostatistics. Article shows the relationship

geostatistics with the scientific picture of the world. It is shown that as the direction of the applied geostatistics provides a mechanism for investigation and management of the education system. Studied aspect of the application of geostatistics in education. It is shown that geostatistics creates information fields that bind together the disparate spatial objects of the same quality. especially geostatistics. It is displaying geostatistics in education reflects the trend of education informatization.

Keywords: education, geoinformatics, statistics, geostatistics, simulation, visualization, geodata, the principles of geostatistics, education management.

УДК 528:004.528

СОСТОЯНИЕ И РАЗВИТИЕ ИНФРАСТРУКТУРЫ ПРОСТРАНСТВЕННЫХ ДАННЫХ

*Василий Тимофеевич Матчин, старший преподаватель,
E-mail: matchin.v@gmail.com,*

*Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Московский государственный технический университет радиотехники, электроники и автоматики,
<https://www.mirea.ru>*

Статья описывает состояние и развитие инфраструктуры пространственных данных (ИПД), как нового универсального информационного ресурса. Показаны недостатки современного состояния использования пространственных данных. Дан лингвистический анализ термина. Отмечен зарубежный опыт создания ИПД. Показаны основные цели и задачи развития концепции по созданию национальной инфраструктуры пространственных данных. Раскрыта организация национальной инфраструктуры пространственных данных. Показаны преимущества, обусловленные созданием и применением национальной инфраструктуры пространственных данных.

Ключевые слова: пространственные данные, инфраструктура пространственных данных, фонды пространственной информации, информационный сервис, базы данных, стандарты пространственной информации.

Введение

В настоящее время актуальными являются вопросы оперативного предоставления пространственной информации потребителям (органам государственной власти, местного самоуправления, организациям и гражданам) и создания условий для межведомственного и межуровневого информационного обмена. Однако большой объем данных, накопленный в результате производственной деятельности различными предприятиями, а также разнообразие форматов, систем координат и технологий затрудняют, а иногда и вовсе препятствуют процессам обмена информацией. Новые требования рынка, предъявляемые к информации о местности, и развитие информационных технологий обуславливают необходимость поиска новых решений на основе методов информатики и геоинформатики [1]. Выход из сложившейся ситуации видится в создании условий, обеспечивающих доступ потребителей к пространственным данным в электронном виде и их эффективное использование. За рубежом подобного рода системы получили название «инфраструктуры пространственных данных» (ИПД). Для решения задач создания ИПД в 2006 году была



В. Т. Матчин