

УДК 316.004

**ОСНОВНЫЕ НАПРАВЛЕНИЯ ВНЕДРЕНИЯ МЕТОДОВ  
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В СОЦИОЛОГИЧЕСКИЕ  
ИССЛЕДОВАНИЯ**

**Кислова Ольга Николаевна** – кандидат социологических наук, доцент кафедры методов социологических исследований Харьковского национального университета имени В.Н. Каразина

*Статья посвящена анализу проблем, связанных с методологическими инновациями в социологии, обусловленными изменением стандартов научной деятельности, произошедшим в связи с бурным развитием и повсеместным внедрением компьютерных технологий. Показано, что компьютеризация способов научного познания не обошла стороной социологию в целом и социологические исследования, в частности. Выделены основные направления внедрения методов интеллектуального анализа данных в практику социологических исследований: анализ результатов количественных исследований; онлайн-исследования; вторичные исследования; анализ социальных сетей; социальный компьютеринг.*

**Ключевые слова:** интеллектуальный анализ данных, компьютеринг, методологические инновации, социологические исследования.

*Стаття присвячена аналізу проблем, пов'язаних з методологічними інноваціями в соціології, які обумовлені зміною стандартів наукової діяльності, що відбулися у зв'язку з бурхливим розвитком і повсюдним впровадженням комп'ютерних технологій. Показано, що комп'ютеризація способів наукового пізнання не обійшла стороною соціологію в цілому й соціологічні дослідження, зокрема. Виділено основні напрямки впровадження методів інтелектуального аналізу даних у практику соціологічних досліджень: аналіз результатів кількісних досліджень; онлайн-дослідження; вторинні дослідження; аналіз соціальних мереж; соціальний комп'ютинг.*

**Ключові слова:** інтелектуальний аналіз даних, комп'ютинг, методологічні інновації, соціологічні дослідження.

*This article deal with the problems associated with the methodological innovations in sociology, due to changes in standards of scientific activity that occurred in connection with the rapid development and widespread introduction of computer technology. It is shown that computerized methods of scientific knowledge are not spared in general sociology and sociological research in particular. The basic directions of the introduction of data mining techniques in the practice of sociological research: an analysis of the results of quantitative research, online research, secondary research, and analysis of social networks, social computing are marked.*

**Keywords:** intelligent data analysis, data mining, computing, methodological innovations, sociological research.

Бурное развитие информационных технологий и, как следствие, расширение возможностей компьютерной обработки информации способствуют появлению множества нетрадиционных подходов к получению нового научного знания, среди которых все большее значение приобретает интеллектуальный анализ данных (ИАД), основанный на использовании методов компьютерного анализа и искусственного интеллекта в поиске латентных знаний, скрытых в массивах исходной информации. Сегодня интеллектуальный анализ данных постепенно становится неотъемлемой частью большинства информационных технологий, проникает во все большее число областей научной и практической деятельности. Обсуждению возникающих в этом контексте вопросов мы посвятили предыдущую статью, где эксплицировали сущность компьютерной эпистемологии, явившейся откликом на стремительную экспансию компьютерных технологий в сферу научного познания, а также пунктирно наметили (но не вполне конкретизировали) взаимообусловленность развития интеллектуального анализа данных и *актуализации* методологических инноваций в современной социологии [1]. Однако данная проблематика настолько обширна, что поднятые вопросы не могли быть (и не были) разрешены полностью. Это и обуславливает наш выбор в качестве *цели* данной публикации выделение основных направлений внедрения методов интеллектуального анализа данных в практику социологических исследований.

Поскольку интеллектуальный анализ данных представляет собой достаточно новую концепцию работы с информацией (в частности социологической), то проникновение его идей в социологические исследования, безусловно, сопровождается активизацией методологических и методических инноваций.

Выбору обозначенной цели способствовало также то, что сквозь призму методологических инноваций всю историю науки можно рассматривать как движение от одного кардинального изменения в развитии методов к другому. Очевидно, что любые методологические инновации являются следствием разработки новых инструментов исследования, способных привести к смене стандартов научной работы и/или появлению новых областей знания (об этом в свое время много писали, например, Н. И. Кузнецова, М. А. Розов, С. Е. Крючкова) [2]. Интеллектуальные компьютерные технологии, появившиеся в конце XX-го – начале XXI-го столетий, являются теми инструментами, которые спровоцировали необходимость ревизии способов научного познания на современном этапе развития науки. Появление компьютерных технологий послужило толчком к смене стандартов научной деятельности и появлению новых специфических областей знания, обозначаемых в англоязычной литературе приставкой "computational": computational philosophy, computational sociology, computational political science, computational linguistics, computational chemistry, computational psychology и др. Социология также оказалась вовлеченной в этот процесс, затронувший всю науку. В частности, сегодня обсуждаются возможности и потенциальные пути развития компьютерной социологии (computational sociology) [3], под которой мы понимаем отрасль научного знания, использующую компьютеринг, т. е. разнообразные компьютерные технологии работы с данными. Отметим, что английское слово "computing", долгое время бывшее синонимом слов "counting" и "calculating", переводимых как "вычисления", в конце XX-го века приобрело иной смысл, вобрав в себя разнообразные операции по использованию компьютеров, электронные процессы, происходящие внутри них (аппаратное обеспечение), управление ними (программное обеспечение) и концептуальные основы (computer science, компьютерные науки). Сегодня слово "computing" используется в смысле "компьютерный, относящийся к работе на компьютере", оно часто рассматривается как синоним "computational", а под "computational science" подразумевают науки, использующие вычисления, выполняемые на компьютере. При этом следует подчеркнуть, что к настоящему времени компьютеры перестали быть просто вычислительными устройствами: они превратились в универсальные инструменты для работы с данными, а слово "компьютеринг" (в украинском варианте – "комп'ютеринг") окончательно интегрировалось как в русский, так и в украинский языки.

Основной наш тезис состоит в следующем: компьютеризация способов научного познания не может обойти стороной социологию в целом и социологические исследования, в частности, она обуславливает поиск новых методологических оснований, адекватных социологии XXI-го века – века зарождения e-НОМО, то есть человека, оснащенного множеством разнообразных "умных" технологий и включенного в IT-среду новых медиа.

Сегодня мы уже наблюдаем некоторые последствия экспансии компьютерных технологий в социологию:

- распространение социологических онлайн-исследований;
- расширение арсенала методов, применяемых для анализа социологических данных, за счет интеллектуальных технологий обработки массивов эмпирической информации (как первичной, так и вторичной);
- становление новых областей социологии, развитие которых было бы невозможным без развития компьютерных технологий (самым ярким примером является Интернет-социология);
- появление концепции социального компьютеринга, под которым понимают новый вид производства знаний, основанный на идеях коллективного разума, "мудрости толпы"<sup>1</sup> и сматрмоба<sup>2</sup>.

Обратимся к выяснению причин, по которым из разнообразия современных компьютерных технологий мы выделяем именно интеллектуальный анализ данных, рассматривая его в качестве фактора инноваций в социологических исследованиях. Естественно, может возникнуть вопрос: "Почему не Интернет или технологии моделирования виртуальных миров?". Потому, что интеллектуальный анализ

<sup>1</sup> "Мудрость толпы" - термин, введенный в научный дискурс Джеймсом Шуровески в книге с одноименным названием [4]. Сегодня "мудрость толпы" является одним из важнейших понятий веба, представляющим значительную ценность для исследований социальных СМИ.

<sup>2</sup> Смартмоб (smart mob – умная толпа) – форма самоструктурирующейся социальной организации, возникающей за счет эффективного использования коммуникационных технологий и повсеместного компьютеринга. Концепция "умной толпы" была предложена в 2002 году Говардом Рейнгольдом, социологом, специализирующимся на исследовании культурных, социальных и политических импликаций в медиа-сфере современности (см. [5]).

данных вездесущ, он является одной из разновидностей *повсеместного компьютеринга*<sup>3</sup>, постепенно проникающего во все сферы человеческой жизни, скромно и незаметно помогающего в ежедневной рутине – дома, в офисе, в дороге ... и, естественно, в научных изысканиях. Так, например, во все современные поисковые системы Интернет внедрены методы интеллектуального анализа. Сканируя и распознавая необходимые для работы тексты, мы, совершенно не задумываясь об этом, используем интеллектуальный анализ. Моделируя виртуальные миры, программисты встраивают в них методы интеллектуального анализа данных, позволяющие искусственным персонажам вести себя "разумно" (см., например: [8]). Кроме того, интеллектуальные методы сегодня включены во все наиболее известные пакеты статистического анализа данных (в частности, в SPSS и STATISTICA), что способствует их применению в обработке результатов социологических опросов наряду с традиционными статистическими процедурами.

Технологии интеллектуального анализа данных относятся к так называемым исчезающим (disappear) технологиям. Причем слово "исчезающие" употребляется для обозначения отнюдь не "вымирающих", отживших свой век технологий, а технологий, которые по мере своего развития становятся все более и более востребованными и привычными, превращаясь в неотъемлемый атрибут жизни и деятельности человека, т. е. исчезают из поля зрения, растворяясь в обыденных практиках. Основоположник повсеместного компьютеринга Марк Уейзер по этому поводу писал: "Самые сильные, совершенные и глубокие технологии – это те, которые исчезают в окружающей нас среде. Они вплетаются в ткань повседневной жизни и становятся неразличимыми на ее фоне" [7, p. 94].

Анализируя влияние повсеместного внедрения методов интеллектуального анализа данных на изменения, происходящие в практике социологических исследований, мы будем выделять области, куда уже начал проникать ИАД, и рассматривать его инновационный потенциал в их дальнейшем развитии.

Наиболее естественной и очевидной сферой применения интеллектуального анализа данных в социологических исследованиях является **расширение возможностей обработки и анализа результатов массовых опросов**. Об этом уже достаточно много написано. Методы интеллектуального анализа данных медленно, но уверенно занимают достойное место в арсенале аналитических техник социолога, однако это направление внедрения методов интеллектуального анализа данных не предполагает изменения методологии социологических исследований. Интеллектуальный анализ здесь просто является очередным звеном в эволюции методов анализа результатов количественных исследований. Он открывает новые горизонты выявления латентных закономерностей, скрытых в массивах социологической информации, но не провоцирует изменений в методологии, применяемой социологами со времен П. Лазарсфельда.

**ИАД в контексте онлайн-исследований: потенциал внедрения data mining и web mining.** Доля онлайн-исследований среди общего числа социологических исследований пока еще невелика, но она неуклонно возрастает, несмотря на определенные проблемы, связанные прежде всего с репрезентативностью получаемой информации. Уже нет сомнений в том, что онлайн-овые социологические исследования превращаются в необходимый атрибут социологии. Поскольку интернетизация захватывает все более широкие слои населения, то можно предположить, что "портрет" Интернет-пользователя в обозримом будущем превратится в репрезентативное отражение "среднего человека" XXI-го века. Что касается методов интеллектуального анализа данных, то их использование в онлайн-овых опросах могло бы дать хорошие результаты не только на этапе обработки полученной информации (где в настоящее время уже используются возможности интеллектуального анализа количественных данных - data mining), но и во время сбора данных.

Сегодня активно муссируется проблематика web usage mining, охватывающая методы сбора и анализа данных об индивидуальных предпочтениях, убеждениях, поведении людей в сети Интернет, а также автоматизированного контент-анализа публичных и частных документов. В силу того, что методы web usage mining имеют значительный исследовательский потенциал, они стали все чаще применяться для сбора персонализированной информации о пользователях Интернет. Проблема состоит в том, что такой сбор информации часто преследует коммерческие цели и временами проводится без информирования и получения согласия исследуемых. Такие "шпионские" методы интеллектуального анализа (использующие специальные "тройные" программы, дающие возможность "наблюдать" за поведением пользователей в Сети и красть личную информацию) в англоязычной литературе часто называют "data snooping" и "data

<sup>3</sup> Повсеместный компьютеринг (ubiquitous computing) – термин, предложенный в 1988 году Марком Уейзером [6; 7] для обозначения результатов внедрения компьютеров и прочих инновационных технологий в повседневную жизнь за счет оснащения обыденных вещей и рабочих инструментов элементами искусственного интеллекта. В качестве примеров можно вспомнить "умный и услужливый" Word, который автоматически исправляет грамматические ошибки в набранном тексте, или "охлаждающую" футболку, которая самостоятельно освежает владельца в случае его повышенной активности.

fishing". Их применение несет угрозу личным свободам человека и влечет за собой необходимость широкого обсуждения моральных аспектов подобных исследований, их соответствия профессиональной этике социолога. Безусловно, в этой сфере сегодня имеется больше вопросов, нежели ответов. Названные методы уже существуют, и от нас зависит, как они будут использованы.

Реалии XXI-го века побуждают современных социологов активизировать работу по расширению арсенала методов социологических исследований, в том числе и за счет включения в него web usage mining. Так, И. Ф. Девятко, отмечая несколько замедленную реакцию специалистов по методологии социальных наук на быстрые изменения способов и технологий репрезентации социальной информации, призывает социологов к более активной деятельности в данном направлении: "Для того, чтобы сохранить способность к лидерству, социологическая методология должна ускорить продвижение в разработке новых методов анализа агрегированной онлайн-информации, новых подходов к контент-анализу символических репрезентаций, новых моделей, описывающих распространение социального влияния и трансляцию мнений и предпочтений во Всемирной сети, сопоставимых с теми блестящими результатами, которые были получены более полувека назад ее отцами-основателями" [9, с. 29].

**Развитие банков социологической информации и ИАД: целесообразность применения идей эксплораторного анализа и инструментов data mining во вторичных исследованиях.** Широкий доступ к результатам исследований, появившийся в связи с развитием банков социологической информации, создает предпосылки проведения вторичных исследований. Актуализация применения методов интеллектуального анализа данных (особенно data mining) во вторичных исследованиях обусловлена необходимостью выявления глубоко спрятанных тенденций, для чего обычно применяются разнообразные "сложные" методы анализа данных. Их применение во вторичных исследованиях особо актуально, поскольку все явные закономерности, которые становятся очевидными уже на уровне дескриптивного анализа одномерных и двумерных распределений, обычно бывают описаны авторами проектов. Однако после публикации основных результатов всегда остается масса побочной информации, исследование которой способно дать априорно не прогнозируемые результаты. Получение такой информации обычно не планируется на уровне разработки программы социологического исследования, а закладываемые в инструментарий индикаторы служат иным целям. Поэтому для анализа подобных дополнительных данных, естественно, не подходят методы традиционной статистики, предполагающей проверку априорных гипотез, что и обуславливает необходимость обращения к методам data mining, имеющим поисковый характер. Однако в практике отечественных аналитиков применение этих методов сталкивается с определенными трудностями, связанными, в первую очередь, с тем, что применение data mining без знания их математических основ весьма проблематично. Именно это и является фактором, тормозящим широкое внедрение data mining в повседневную практику социологического анализа, в том числе и во вторичные исследования.

**Анализ социальных сетей и развитие social network mining.** Кто бы мог подумать лет 10 назад, что социальные сети приобретут такую актуальность в жизни современного человека и станут предметом пристального внимания социологов?! Исследования социальных сетей сегодня актуализируются в связи с тем, что подавляющее число пользователей Интернет, особенно молодых людей, буквально "живут" в них, виртуальное общение в сетях приобретает все большее значение и даже иногда заменяет общение с реальными людьми. В качестве подтверждения данного тезиса мы не будем приводить ссылки на многочисленные исследования, касающиеся социальных сетей, а обратимся к данным социологического опроса первокурсников Харьковского национального университета имени В. Н. Каразина, проведенного в конце 2011 года<sup>4</sup>. Опрос показал, что практически все первокурсники пользуются Интернетом (99 %), 95 % из них зарегистрированы в социальных сетях, причем 67 % – в нескольких одновременно. Ответы на вопрос "Как часто вы проводите время в социальных сетях?" распределились следующим образом: 3-4 раза в месяц - 7 %; 3-4 раза в неделю - 21 %; практически каждый день - 44 %; ежедневно несколько раз - 28 %. Данное распределение не нуждается в пояснениях, важность виртуального общения в жизни наших студентов очевидна.

В настоящее время наблюдается активная экспансия методов интеллектуального анализа данных в исследования социальных сетей [10], результатом которой явилось формирование нового направления социологических исследований, получившего название "social network mining". Данные, которые используются для этого вида интеллектуального анализа, могут быть числовыми, текстовыми, а также

<sup>4</sup> Речь идет об исследовании, проведенном в период с 24 октября по 8 ноября 2011 года сотрудниками научно-исследовательского Института социально-гуманитарных исследований Харьковского национального университета имени В.Н.Каразина совместно с социологическим факультетом этого же ВУЗа. Опрошены все первокурсники (за исключением отсутствующих и иностранных студентов), N = 1002.

мультимедійними. Начиная с 2007 года, ежегодно проводится международный симпозиум "Social Network Mining and Analysis" [11], в январе 2011 года начал издаваться журнал "Social Network Analysis and Mining" [12], в котором публикуются статьи, посвященные как разработке новых инструментальных средств анализа социальных сетей, так и специфике их использования в исследовании конкретных социальных феноменов.

Под влиянием роста числа приложений социальных сетей возникла новая парадигма компьютеринга – **социальный компьютеринг**, под которым понимают "новый вид нематериального производства (производства знаний), включающий в себя новую среду производства, нового субъекта производства и новый способ производства знаний" [13]. Социальный компьютеринг предполагает изучение социального поведения с использованием современных компьютерных систем, он интегрирует в себе многообразие методов познания "мудрости толпы", объединяя исследования социальных сетей с методами предсказания социального поведения, динамики культур и т. п. [14]. Интеллектуальный анализ данных в подобных исследованиях является необходимым компонентом, настолько прочно вплетенным в ткань этих исследований, что его роль даже не обсуждается.

Начиная с 2008 года, проводится ежегодная международная конференция "Social Computing, Behavioral Modeling, Prediction" [14], действуют специализированные лаборатории, исследующие возможности использования социального компьютеринга для предсказания будущих событий [15]; в России в июне 2012 в МГГУ им. М.А.Шолохова планируется проведение первой международной конференции "Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты". Все это свидетельствует о пристальном внимании научного сообщества к проблемам, возникающим в процессе трансформации Homo Sapiens в e-НОМО – человека нового тысячелетия, оснащенного многочисленными интеллектуальными технологиями (в руке смартфон, под мышкой ноутбук, в ушах наушники ...). Естественно, что социология не может не рефлексировать по этому поводу, не может остаться в стороне от социальных изменений, обусловленных вторжением технологических новшеств в жизнь людей.

Какова новая, "возникающая" на основе повсеместной компьютеризации социальная реальность? Этот вопрос сегодня волнует социологов так же, как и представителей иных наук. С одной стороны, включение в глобальную электронную цивилизацию, несущую новые возможности образования, коммуникации, личного развития и развлечений, с другой – растущая зависимость от компьютеризированной среды обитания e-НОМО, опасность тотального контроля, потенциальный уровень которого превосходит все ранее существовавшие авторитарные системы. Многоплановое взаимодействие названных векторов развития предполагает разнообразные нелинейные эффекты, поэтому сегодня сложно однозначно ответить на поставленный вопрос. Однако с полной определенностью можно утверждать, что методы исследования этой "возникающей" социальной реальности должны быть адекватны ей. Именно поэтому сегодня многие социологи говорят о назревшей необходимости изменений методологии социальных исследований, сравнивают настоящее положение с ситуацией, сложившейся в 1930–1950-е годы, когда революционные изменения в средствах массовой коммуникации спровоцировали П. Лазарфельда и его коллег заняться изучением социальных эффектов новых (для того времени) медиа, что привело к формированию методологического "ядра" социологии (см. [9]). Сегодня появились другие медиа (прежде всего Интернет), соответственно, вновь возникла потребность пересмотреть методологический аппарат социологии. Этот пересмотр уже начался, ведутся дискуссии относительно перспектив и опасностей использования многообразной информации о группах и индивидах, разбросанной во Всемирной сети, о методах, которые могут быть использованы социологами для получения и анализа этой информации. В этом контексте интеллектуальный анализ данных (как концепция и как отдельные методы) привлекает все большее внимание, становится основой определенных методологических инноваций. Это внимание, как мы полагаем, постепенно сойдет на нет, когда интеллектуальный анализ данных "растворится" в социологической специфике.

В качестве заключения подчеркнем, что мы, естественно, не можем претендовать на полноту описания методологических инноваций в социологии, возникающих в связи с компьютеризацией способов научного познания. Мы лишь обозначили те из них, которые сегодня являются наиболее актуальными и где интеллектуальный анализ данных в ближайшее время будет максимально востребован.

**Литература:**

1. Кислова О. Н. Интеллектуальный анализ данных и компьютерная эпистемология как факторы методологических инноваций в современной социологии / Кислова Ольга Николаевна // *Методологія, теорія та практика соціологічного аналізу сучасного суспільства. Збірник наукових праць*. – Харків: видавничий центр Харківського національного університету імені В. Н. Каразіна, 2011. – С. 241-247.
2. Кузнецова Н. И. О разнообразии научных революций / Н. И. Кузнецова, М. А. Розов // *Традиции и революции в истории науки*. – М.: Наука, 1991. – С. 60–83; Крючкова С. Е. *Инновации: философско-методологический анализ* / С. Е. Крючкова. – М.: "Станкин", 2001. – 186 с.
3. Кислова О. Н. Компьютерная социология: генезис и перспективы развития / Кислова Ольга Николаевна // *Вісник Харківського національного університету імені В. Н. Каразіна "Соціологічні дослідження сучасного суспільства: методологія, теорія, методи"*. – № 941. – 2011. – С. 20-28.
4. Шуровески Дж. *Мудрость толпы. Почему вместе мы умнее, чем поодиночке, и как коллективный разум формирует бизнес, экономику, общество и государство* / Джеймс Шуровески : Пер. с англ. - М.: Вильямс, 2007. – 304 с:
5. Рейнгольд Г. *Умная толпа: Новая социальная революция* / Говард Рейнгольд : Пер. с англ. – М.: ФАИР-ПРЕСС, 2006. — 416 с.
6. Weiser M. D. Ubiquitous Computing [Electronic resource] / Mark Weiser // Weiser web page. – Available at: <http://www.ubiq.com/hypertext/weiser/UbiHome.html>
7. Weiser M. D. The Computer for the 21st Century [Electronic resource] / Mark Weiser // *Scientific American*. – 1991. – № 265(3). – P. 94-104. – Available at: <http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>
8. Шампадар А. *Искусственный интеллект в компьютерных играх. Как обучить виртуальные персонажи реагировать на внешние воздействия* / Алекс Дж. Шампадар. – М.: Вильямс, 2007. – 768 с.
9. Девятко И. Ф. *Онлайн исследования и методология социальных наук: новые горизонты, новые (и не столь новые) трудности* / Девятко Инна Феликсовна // *Онлайн-исследования в России 2.0*. – М.: РИЦ "Северо-Восток", 2010. – С. 17-30.
10. Jensen D., Neville J. *Data Mining in Social Networks* [Electronic resource] / David Jensen and Jennifer Neville // *Dynamic Social Network Modeling and Analysis*. – Washington: National Academy Press, 2002. – Available at: <http://www.cs.purdue.edu/homes/neville/papers/jensen-neville-nas2002.pdf>; Srivastava J. *Data Mining for Social Network Analysis* [Electronic resource] / Jaideep Srivastava // *Australasian Data Mining Conference (AusDM - 2007)*. – Available at: <http://www.docstoc.com/docs/40260549/Data-Mining-for-Social-Network-Analysis>
11. *Social Network Mining and Analysis* [Electronic resource]. – Available at: : <http://www.snakdd.com/>;
12. *International Journal Social Network Analysis and Mining* [Electronic resource]. – Available at: [http://www.springerlink.com/content/1869-5450/preprint/?sort=p\\_OnlineDate&sortorder=desc](http://www.springerlink.com/content/1869-5450/preprint/?sort=p_OnlineDate&sortorder=desc)
13. Карелов С. В. *Социальные платформы и социальный компьютеринг* [Электронный ресурс] / Сергей Карелов // *iToday.ru* - онлайн-газета о Сети и IT. – Режим доступа: <http://www.itoday.ru/blog/karelov/301.html>
14. *Social Computing, Behavioral Modeling, Prediction: Home of the SBP Conferences* [Electronic resource]. – Available at: <http://sbp.asu.edu/>
15. *HP Social Computing Laboratory* [Electronic resource]. – Available at: [http://www.hpl.hp.com/research/idl/?jumpid=reg\\_R1002\\_USEN](http://www.hpl.hp.com/research/idl/?jumpid=reg_R1002_USEN)