

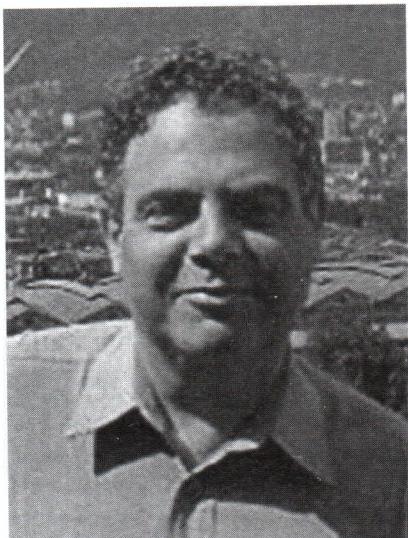
Computer Science борется за справедливую оценку

В продолжение темы мировых рейтингов университетов, о которых наш журнал писал в № 15, 2010 г. («Как попасть в топ-100 мировых рейтинговых вузов») и в № 5, 2011 г. («СПбГУ в контексте мировых университетских рейтингов»), мы продолжаем публиковать материалы, связанные с современными проблемами рейтингов вузов и отдельных наук.

В журналах Nature и Science в принципе не публикуются статьи посвященные программированию, поэтому программисты заведомо попадают в невыгодную ситуацию по сравнению с физиками, химиками, биологами и другими представителями естественных наук. В то же время данная проблема волнует не только европейских ученых, работающих в области Computer Science (CS), но и представителей ряда других наук.

Журнал «Санкт-Петербургский университет» предлагает вниманию читателей материал по про-

блеме оценки результатов исследований в этой области, который подготовили четыре известных в Computer Science европейских ученых: профессор **Берtrand МЕЙЕР** (Bertrand Meyer), заведующий кафедрой в Университете ETH (Цюрих, Швейцария), президент ассоциации Informatics Europe; профессор **Кристина ЧОППИ** (Christine Choppuy), Университет Париж 13 (Франция), вице-президент ассоциации Informatics Europe; профессор **Ян Ван ЛИВЕН** (Jan van Leeuwen), Университет Уtrecht (Голландия), вице-президент Informatics Europe;



Берtrand МЕЙЕР



Кристина ЧОППИ

профессор **Йорген СТАУНСТРУП** (Jorgen Staunstrup), проректор ИТ-университета Копенгагена (Дания).

Материал предоставлен заведующим кафедрой системного программирования СПбГУ, доктором физико-математических наук, профессором А.Н. Тереховым, публикуется в переводе и с разрешения авторов.

INFORMATICS EUROPA —

Ассоциация факультетов информационных технологий европейских университетов, основанная в 2005 году.

<http://www.informatics-europe.org/>

I. ОЦЕНКА ИССЛЕДОВАНИЙ И ЕЁ РОЛЬ

Исследование как таковое несет в себе соревновательное начало. Исследователи привыкли к тому, что их труд постоянно оценивается. Любая представленная для публикации статья (а порой даже и приглашенная) вначале оценивается коллегами, причем очень часто работа отклоняется. Даже если автор — маститый ученый, каждая его новая статья будет проходить экспертное оценивание. Будучи однажды опубликованной, статья исследователя вновь и вновь будет сравниваться с работами других ученых. Кроме того, что работы ученых оцениваются, им самим приходится постоянно выступать судьями при рецензировании статей, представленных к публикации, (кроме, пожалуй, самых молодых исследователей), участвовать в работе различных комитетов, оценивать заявки для участия в конкурсах различных фондов, отвечать на вузовские запросы по оценке исследовательских проектов.



Йорген СТАУНСТРУП

Вся система взглядов в исследовательской среде опирается на оценки, которые одни ученые выставляют другим. Критерии должны быть справедливыми (насколько вообще справедливой может быть любая деятельность, связанная с человеческими суждениями), понятными и одобренными целевой научной аудиторией. Несмотря на то что подобное оценивание проводят в разных дисциплинах, критерии из одной области не должны навязываться для оценки в других (например, когда критерии из науки с более длинной, устоявшейся историей переносятся на молодую науку, развивающуюся по своим принципам).

1. COMPUTER SCIENCE — это отдельная дисциплина, сочетающая черты науки и инженерии. Оценивание исследований в CS должно соответствовать специфике дисциплины.

2. Отличительной чертой публикаций в области CS является важность конференций, часть из которых имеют жесткий отбор, а также монографий. Публикации в журналах, хотя и необходимы для развернутого описания отдельных тем, не могут оцениваться выше, чем главные конференции или монографии.

3. Результатом значительной части исследований в CS является не публикация, а чаще всего программный продукт. Если оценивать научное влияние, то такие результаты могут быть не менее важны, чем публикации.

4. В культуре публикаций в области CS порядок, в котором указываются авторы публика-



Ян Ван ЛИВЕН

ций, не зависит от их вклада в статью. Поэтому, если нет специального примечания, порядок не может служить критерием оценивания.

5. Любые показатели количества публикаций какого-либо из авторов (абсолютные или относительные) не могут быть использованы в качестве индикаторов научной ценности. Они измеряют такую сторону исследовательской работы, как продуктивность в написании статей, а не их качество или вклад в науку.

6. Влияние на науку, выраженное в числах (например, количество цитирований), может иметь место, но не должно быть единственным мерилом для оценки. Любое использование подобной методики должно быть индивидуальным и проводиться с учетом экспертной оценки, дабы избежать множества ошибок. Кроме того, как было сказано выше, необходимо оценивать не только публикации, но и другие результаты исследований в CS.

7. Любой критерий оценки, особенно если он выражается конечным числом, должен основываться на простых и доступных для проверки принципах.

8. Численные индикаторы не должны использоваться для сравнения ученых или научных трудов в разных дисциплинах.

9. База данных ISI Web of Science не является адекватным источником информации для оценки количества и цитируемости публикаций в большей части дисциплин CS и потому не должна применяться в этих целях. Желательными альтернативами, пусть и не идеальными, являются, например, Google Scholar, CiteSeer и (потенциально) Цифровая библиотека ACM.

10. Критерии оценки сами должны быть объектом оценки и возможного пересмотра.

сих пор ученые в области CS не предприняли значительных шагов для объяснения принципов и отличительных черт своей науки ученым из других областей.

Одной из основных задач Ассоциации Informatics Europe является широкое освещение специфики этой науки и требований к ее оцениванию.

Данный материал, подготовленный Informatics Europe, основывается на результатах работы CRA и освещает лишь некоторые специфические для Европы проблемы, такие как языковое разнообразие, по простой причине — критерии для определения качества научного исследования вряд ли отличаются во всем мире. Некоторые вопросы, затронутые в материале CRA, рассматриваются более подробно, учитываются также и те изменения, которые произошли с нашей наукой после 1999 года.

Дисциплина CS, в которой делается основной упор на анализ, является ядром информационных технологий, но не единственной их ставной частью. Поэтому, хотя многие наши выводы подходят и для других ИТ-дисциплин, их следует адаптировать для отдельных дисциплин, таких как, например, цифровые СМИ, так как они могут пересекаться с гуманитарными науками, имеющими другие традиции публикаций результатов исследований.

II. НАУКА COMPUTER SCIENCE И ЕЁ РАЗНООБРАЗИЕ

Сама по себе наука Computer Science связана не столько с компьютерами, сколько с обработкой информации на них с использованием алгоритмических техник. Распространенный в Европе термин «информатика» имеет более широкий смысл, включая в себя дисциплины, связанные с человеческим участием в информационных технологиях. В нашем отчете мы будем рассматривать науку именно с этих позиций, не делая различий между двумя терминами.

Исследования в области CS делятся на три большие категории: «Теория», «Системы» и «Приложения». Это деление не является абсолютным, так как исследования в одной из областей могут пересекаться с другими, однако удобно для общей классификации.

Теоретические исследования

связаны с концептуальными подходами к пониманию вычислений, алгоритмов, структур данных и других аспектов компьютерных приложений. Их можно также разделить на три подкатегории:

- алгоритмы, вычислительная сложность и комбинаторика (математические модели для понимания машин и вычислений);
- семантика, спецификации и доказательства (математические модели программирования и языков программирования, в основном для гарантии правильного функционирования);
- вычислительная наука (математические модели для высокопроизводительных вычислений).

Все эти три категории активно используют математику, хотя математика, присущая двум первым категориям, в традиционном научном образовании не ставится на первое место: логика, формальные языки, теория автоматов.

Исследования систем посвящены созданию неких объектов и оценке их свойств. Эти объекты могут быть программами, а также системами, которые:

- включают в себя программную часть наряду с другими элементами, как, например, во «встроенных системах» (мобильные телефоны, поезда, авиадиспетчерские комплексы);
- включают в себя как программную, так и аппаратную часть, или в «информационных системах управления»;
- включают в себя программную часть и организационные процессы.

Категория «Системы» включает в себя две основные области для исследований:

■ создание систем — исследует прототипы систем, а также ПО, которое может быть тиражировано;

■ исследование свойств существующих информационных систем и процессов — это так называемая экспериментальная CS. В ней используются некоторые методики, взятые из естественных наук и статистики (для «анализа производительности», например, при изучении пропускной способности сетей).

Эти две области часто объединяются, поскольку ученые, созда-

ющие системы, также анализируют их свойства. По областям приложений Исследования систем можно разделить на системное программирование (изучает лучшие способы создания и сопровождения высококлассных программных систем, в основном больших и сложных), языки программирования и их реализация (компиляторы, интерпретаторы), человеко-машинный интерфейс и графика, базы данных (занимается изучением управления большими объемами информации), организация сетей и операционные системы, безопасность (решает вопросы поддержания целостности и сохранности информации) и другие.

Сущность «Теории» и «Исследования систем» охватывает все области приложения компьютеров. **Исследование Приложений** направлено на изучение специфики приложений компьютерных вычислений в различных дисциплинах, например, вычислительная химия или вычислительные финансы. Эту область часто называют «Computational X». Оценка исследований в этой области требует объединения критериев, присущих CS и тем областям, для которых создаются приложения — так называемая область X. Данный материал ограничивается анализом первых двух областей («Теория» и «Исследование систем», иногда эти две области вместе еще называют «основная информатика»), оставляя за читателем право самому определять критерии оценки исследований в области Приложений, в зависимости от конкретного X.

Теория CS во многом похожа на такие науки, как математика и физика. Исследование систем имеет ряд черт, аналогичных инженерным наукам, таким как электротехника и машиностроение. Такая двойственность, с одной стороны, делает CS привлекательной дисциплиной, с другой стороны, обязывает более аккуратно подходить к оцениванию ее результатов, так как требует найти нужный набор критериев, присущих как чистой науке, так и инженерным исследованиям.

Исследование в области CS, будь оно более ориентировано на Теорию либо на Исследование систем, содержит строгий набор характеристик, которые сформировались за полувековую исто-

рию этой науки. (Конечно, при желании можно найти корни этой науки еще у древних греков, арабов или индийцев, тем более не следует забывать про пионеров математики, таких как Паскаль. Однако наукой как таковой CS стала только после Второй мировой войны, с появлением компьютеров, а первые кафедры по этому направлению появились в университетах в 60-е годы прошлого века.) Воплощением духа дисциплины является набор основополагающих концепций и парадигм, таких как понятие алгоритма, вычислимость, вычислительная сложность, двойственность спецификации/реализации, рекурсия, масштабируемость, двойственность функции/данных, параллелизм, понятие алгоритмической сложности, двойственность постоянных/переменных, моделирование, понятие протокола, трансляция между языками, взаимодействие с пользователем и так далее. Понимание всего этого набора концептуальных основ необходимо любому ученому в области CS, тогда как не все ученые из других дисциплин осведомлены даже о существовании этого массива объектов для изучения, считая компьютер просто инструментом, а компьютерные вычисления вспомогательной задачей, а не объектом научного исследования. Такое мнение тем более превратно, что ускорение развития других наук стало возможным во многом благодаря исследованиям в CS, которые позволили другим дисциплинам совершать невозможные до того действия — от масштабных вычислений до использования глобальных источников знаний в Интернете. Ученые в области CS несут ответственность за то, чтобы донести до своих коллег как значимость своей науки, так и ее специфические черты.

На основании вышеизложенного можно вывести первую рекомендацию:

1. Computer Science — это самобытная дисциплина, объединяющая в себе науку и инженерию. Эту особенность следует учитывать, оценивая ученых в этой дисциплине.

III. КУЛЬТУРА ИССЛЕДОВАНИЙ В ОБЛАСТИ CS

Исследования в CS имеют определенные отличительные свойства, которые появились то ли из самой природы этой науки, то ли по каким-либо другим обстоятельствам.

Каналы публикаций

Особенность культуры публикаций по теме CS такова, что презентация результатов исследований чаще всего проходит на какой-нибудь авторитетной конференции — в отличие от дисциплин, где законченное исследование впервые публикуется в специализированном журнале, а на конференциях идет лишь обмен мнениями относительно представленных (часто еще сырьих) результатов. Значительная часть публикаций в CS печатается в Сборниках трудов конференций, таких как POPL, PLDI, OOPSLA, ICSE, SIGGRAPH, Eurographics и других. На отдельные CS конференции в результате отбора допускаются только от 10% до 20% поданных на рассмотрение работ. Например, в 2007–2008 году процент одобренных к участию заявок был такой:

- ICSE (International Conference on Software Engineering, системное программирование) — 13%;
- OOPSLA (Object-Oriented Programming, Systems, Languages & Applications, объектно-ориентированное программирование) — 19%;
- POPL (Symposium on Principles of Programming Languages, языки программирования) — 18%.

Журналам в нашей науке отведена своя роль — часто там печатают расширенные версии статей, ранее представленных на конференциях, включающие детали, которые нельзя было уместить в ограниченном формате трудов конференции. И хотя многие ученые находят время на журнальную публикацию таких более расширенных версий, есть множество известных ученых, публикующих свои результаты в трудах конференций. Это может вызвать проблемы в оценивании исследований по сравнению с другими дисциплинами, так как коллеги, занимающиеся ими, могут считать публикации в журналах главным мерилом узнаваемости ученого.

Такая же проблема связана с книгами. Во многих дисциплинах книги не рассматриваются в качестве весомого научного вклада. В CS (так же, как и в некоторых гуманитарных науках) книги, наоборот, могут оказывать наиболее заметное влияние на науку. Например, на вопрос о самой важной публикации в истории CS многие ученые, работающие в этой области, назовут «Искусство программирования» Дональда Кнута. В развитии Системного программирования эпохальные концепции, такие как, скажем, шаблоны проектирования, также впервые были описаны в книгах, а уже потом в других видах публикаций.

Типичный процесс публикации новой идеи состоит из следующих последовательных шагов (в порядке возрастания важности). Вначале автор публикует внутренний отчет в учреждении, в котором работает, где фиксирует сам факт идеи. Затем он может подать статью на обсуждение на каком-либо семинаре. Обычно семинары проводятся в рамках какой-либо большой конференции широкой тематики, и принятие статьи в программу семинара не требует много времени, при этом давая возможность участвовать в конференции. Следующий шаг — это подача статьи на конференцию, и здесь престижность конференции (а степень престижности известна ученым этой дисциплины) показывает уровень работы. Во многих случаях на этом процесс заканчивается, однако автор может доработать статью до журнальной, более длинной версии. А может решиться написать книгу с полным освещением своих идей.

Любой процесс оценки ученого в области CS должен быть согласован с общей культурой публикаций в этой дисциплине.

2. Характерной особенностью публикаций по теме CS является важность статей на конференциях и книг. Статьи в журналах не являются более важными публикациями, чем книги или труды престижных конференций.

При этом, даже если корректно подходить к учету различных

типов публикаций, они не являются единственными результатами научного исследования (в отличие от некоторых других дисциплин). Для некоторых ученых, особенно работающих в области Исследования Систем, часто лучший способ показать значение исследования — это написать программу, создать новый инструмент и т.д., что привлекет внимание коллег и всей мировой общественности. Это может быть более показательным результатом, чем десяток статей по теме. К примеру, успех Google связан с использованием определенного алгоритма (как мы помним, алгоритм — это одна из основ в науке CS) PageRank, определяющего популярность отдельной веб-страницы по числу ссылок на нее, и вычисляемого глобально итеративным алгоритмом. До того как поисковик Google стал коммерческим продуктом, он являлся успешным исследовательским проектом, результатом которого стала статья о PageRank и собственно сам сайт поисковой системы, программное обеспечение, заложенное в его основу. В статье описывался оригинальный алгоритм, один из сотен появляющихся новых алгоритмов. А вот как раз сайт (мы не будем говорить о его последующем впечатляющем коммерческом успехе) показал значение всего исследования, которое не могла бы показать публикация — он продемонстрировал масштабируемость решения. И если бы авторы (С.Брин и Л.Пейдж) остались чистыми учеными и их работа начала бы оцениваться в научных кругах, то программная реализация была бы не менее важна, чем статья.

Однако подобный принцип в оценивании должен быть применен предельно аккуратно — миллион скачиваний программы еще не говорит о ее научной ценности. Публикация же, которую уже оценили специалисты, представляется более удобной единицей измерения при формировании рейтингов. Вот только при оценке работ в CS, особенно в разделе, касающемся Систем (практики), одного подсчета публикаций без учета успешных реализаций идей недостаточно.

3. Для того чтобы оценить влияние исследования (особенно в области Систем), необходимо учитывать не только публикации, но и такие объекты, как программные реализации. Они могут быть не менее важны, чем статьи.

Соавторство

Отдельная тема для обсуждения — оценка вклада отдельного ученого в работу научного коллектива, которую часто определить непросто. В различных дисциплинах приняты свои традиции: в математике часто число соавторов статьи небольшое, а в исследованиях в области экспериментальных дисциплин, наоборот, обычно много участников, каждый из которых хочет, чтобы его участие было не обделено вниманием. Практика публикаций в CS тоже отличается от других дисциплин. Очевидно, что число соавторов статей, связанных с теорией, меньше, чем у статей, связанных с прикладными исследованиями, но при этом в общем случае это число меньше по сравнению с естественными науками и больше, чем у математиков. Рассмотрим это на примере разных изданий за 2007–2008 гг.

■ Журнал «Nature» (за год): наибольшее количество соавторов для статьи — 22, среднее число — 7,3;

■ Журнал «American Mathematical Monthly»: максимум — 6, в среднем — 2;

■ OOSPLA и POPL: максимум — 7 в среднем — 2,7.

Для дисциплин, у которых при написании статьи практикуется участие большого количества соавторов, характерно выстраивание списка авторов в порядке, определяемом их личным вкладом в работу. Это не принято в CS (кроме тех случаев, когда фамилии аспирантов ставятся, как правило, перед фамилией научного руководителя в их совместной статье) — поэтому не так просто определить вклад каждого соавтора; получаем конфликт между общей оценкой значимости статьи и возможностью разделения этой значимости между авторами.

Опасность здесь заключается в том, что оценка может быть исажена, если будут приниматься во внимание те факторы, которые в конкретной дисциплине не имеют

значения. Отсюда следующая рекомендация.

4. Порядок, в котором расставляются фамилии авторов в публикациях в области CS, обычно не имеет значения. В том случае, если это не оговорено явно, порядок фамилий не может служить основанием для сравнения вклада отдельного ученого.

IV. БИБЛИОМЕТРИЯ

Все больше усиливается давление на университеты со стороны государства и общества с требованиями предъявлять результаты научной деятельности. Поэтому администрациям университетов по всему миру приходится использовать в своих отчетах количественные показатели, которые состоят из базовых и производных оценок. Иногда для их определения используют общее понятие — библиометрия.

К базовым оценкам относятся:

- простой подсчет количества публикаций;
- число публикаций с учетом их веса (определяется по некоторым общепринятым критериям престижности отдельных видов публикаций, например, в специализированных журналах и трудах конференций);
- число цитирований, которое измеряет не количество опубликованных статей, а их влияние, определяемое количеством других работ, которые ссылаются на данную публикацию.

Производные оценки, такие как h-индекс (основной или нормализованный) или g-индекс, подсчитываются на основе индексов цитирования методами, описанными ниже.

Идея использовать такие индикаторы для оценки исследований вызвала немало отрицательных отзывов, таких как статья Д.Парнаса, известного ученого в области CS, и коллективное письмо 93 швейцарских профессоров этой дисциплины.

В то же время совсем избавиться от числовых критериев оценки вряд ли возможно. Даже если отвлечься от политического контекста (необходимость университетам показывать осознанные результаты, понятность для восприятия чисел), предлагаемые альтернативы не всегда

лучше. Многие ученые в области CS говорят, что peer review (оценка коллег) является, по их мнению, наилучшим методом для оценки качества работы, но этот метод также имеет свои особенности:

■ оценка работ коллегами сильно зависит от выбора рецензента, его занятости (наиболее авторитетные, как правило, наиболее заняты);

■ если peer review сделать единственным способом оценивания результатов, ученым будет никогда заниматься исследованиями, они будут постоянно загружены оцениванием результатов коллег;

■ по своей природе, оценка коллег может быть предвзятой или необъективной. Работу не может оценить большое количество ученых, обычно речь идет об оценке со стороны нескольких людей — поэтому итоговая оценка сильно зависит от конкретного набора рецензентов.

Выход кроется в комбинации методик peer review и оценки по объективным индикаторам. Последние должны применяться очень внимательно, об этом мы поговорим позже, а также обязательно должны использоваться с учетом анализа со стороны других ученых.

5. Числовые показатели, такие как количество публикаций, не могут быть использованы как единственный инструмент оценки. Они должны быть критически оценены, в частности, чтобы избежать ошибок, и дополнены отзывами специалистов из той же области и перечнем не только публикаций, но и других результатов.

При определении подходящих числовых показателей для сравнения необходимо учесть, что в первую очередь измерять надо качество, а не количество. Подсчет публикаций (общий или взвешенный) отражает лишь внешнюю активность ученого. Если считать их единственным или наиболее главным мерилом, это приведет к обесцениванию публикаций (серезная и актуальная проблема в науке, когда в потоке работ трудно заметить действительно стоящие). Мы получим тонны журналов, которые никто не будет читать, сотни конференций, интересных только докладчикам, и «стахановские» ре-

зюме ученых, где количество будет преобладать над качеством.

6. Число публикаций не является адекватным показателем научного веса. Они отражают лишь продуктивность, а не качество или влияние работы.

Некоторые методики подсчета количества публикаций используют в итоговой оценке такой показатель, как авторитетность издания, в котором опубликована работа, основываясь на определенных рейтингах. Такой подход содержит в себе часть преимуществ метода peer review, так как последний применяется при отборе работ для публикации и релевантен авторитетности издания. Однако есть и недостатки, связанные со справедливым определением рейтинга издания. Чаще всего отдельного ученого оценивают по его публикациям. Многие аналитические агентства имеют заранее составленные рейтинги престижных изданий, которые вызывают многочисленные споры и которые трудно использовать в такой быстроразвивающейся дисциплине, как CS. Рейтинги изданий часто основываются на таком звучном критерии, как импакт-фактор публикаций, который вычисляется автоматически, от чего не становится более достоверным. Даже один из редакторов журнала «Nature» — издания с одним из самых больших импакт-факторов публикаций — высказался негативно об этой концепции. Единственным способом создать действительно вызывающие доверие рейтинги изданий является процесс, в котором будут участвовать эксперты в конкретных дисциплинах. При отсутствии такого консенсуса внутри определенной области науки, определение веса публикации по рейтингу издания никогда не станет существенно лучшим критерием для оценки.

Количество цитирований публикации больше отражает их влияние. Определить этот показатель можно из различных баз данных: ISI Web of Science, которая плохо подходит для CS; CiteSeer, которая привлекала достаточно много внимания, когда была запущена, но сейчас активно не поддерживается; ACM Digital Library, — цифровая библиотека основного сообщества в CS;

Google Scholar — наиболее активно использующийся сейчас ресурс, но основанный на проприетарных (ПО, имеющие собственника) критериях. Показатели цитируемости тоже подвергаются серьезной критике. Основной причиной такой негативной реакции является широкое использование ISI Web of Science для оценки ученых в области CS. Но и использование более достоверных баз данных связано с определенными проблемами.

■ **Фокус.** Нам необходимо оценить качество исследований, при этом качество публикаций — лишь один из показателей качества исследований. Их известность — лишь один из показателей их качества. Количество цитирований — один из показателей известности.

■ **Идентификация.** Неверное произношение или ошибка в записи имени автора может уменьшить количество зафиксированных цитирований. И наоборот — часто работы разных авторов подсчитывают вместе, особенно после того, как в обычную практику вошло использование только инициалов имени, поэтому в таких распространенных сочетаниях как «J Smith», «J Schmidt» или «J Dupont» будут учитывать труды многих ученых. При написании китайских и венгерских имен часто меняют местами имя и фамилию. В зоне особого риска — имена с нестандартными символами, с которыми во всех базах данных (невероятно, но факт) возникают проблемы. Поэтому публикации, скажем, профессора Fröhr могут быть разбиты на «нескольких ученых» — Fröhr, Froehr, Frohr и даже двух соавторов Fr и Hr!

■ **Прочие ошибки.** В отчете института INRIA по библиометрии приводится пример девяти различных способов написания названия института в описании работ всего 4 авторов из INRIA. Другой пример связан с ETH (Цюрих), чьи сотрудники в различных базах данных не просто определены в разные наименования этого вуза, но часто еще и причислены к институту EPFL в Лозанне (иногда и один, и другой вуз в базах данных называют Швейцарский федеральный технологический институт — Swiss Federal Institute of Technology). В связи с этим требуется быть особенно внимательным при использовании баз данных для

сравнения различных институтов – например, если ETH и EPFL сравниваются между собой в целях определения размера финансирования (при этом существует множество известных рейтингов, например, Шанхайский рейтинг, которые сравнивают между собой вузы, опираясь на цифры из баз данных).

■ Язык. Довольно часто в базах данных упор делается на документы на английском языке. В случае с CS это не самая большая проблема, как, например, в математике или гуманитарных дисциплинах.

■ Искажение. Обычно во вступительном разделе статьи делаются ссылки на различные обзоры, чтобы познакомить читателя с тематикой. Поэтому вместо статьи, в которой впервые было представлено какое-то научное открытие и которая обычно тяжела для понимания и прочтения, авторы следующих статей по той же тематике ссылаются не на первую статью, а на более позднюю (часто другого автора), упрощенную для понимания. Парадокс, но ключевая для науки статья, в которой было представлено понятие NP-полноты, менее цитируется, чем последующие статьи, популяризующие эту идею.

■ Неверная интерпретация. Цитирование статьи может подразумевать критику, а не только одобрение. Например, многие авторы, публикующие статьи о своих программах проверки корректности, ссылаются на известную статью, описывающую некий протокол, при этом приводят данные о том, что их инструмент, так же как и остальные, нашел в этом всем известном протоколе всем известную ошибку.

■ Время. Самые свежие публикации по объективным причинам накопили меньшее количество цитирований. Отсюда следует, что справедливую оценку какой-то работе можно дать только по прошествии определенного времени.

■ Размер. Простой подсчет цитат не учитывает разные размеры целевых аудиторий для разных направлений науки. Значимая для истории развития определенной области науки статья может быть процитирована реже, чем проходная работа в «модной», активно обсуждаемой дисциплине.

■ Работа сообща. Несколько авторов образуют «синдикат», перекрестно цитируя друг друга.

■ Лесть. Некоторые авторы, представляющие свои работы на конференцию, ссылаются в них на статьи членов программного комитета, надеясь (бессстыдно) увеличить шансы на одобрение.

Два последних примера наглядно иллюстрируют, как принятые методики оценки оказывают негативное влияние на научную работу в целом. Действительно, если исследователей оценивать по каким-нибудь количественным критериям, они будут вынуждены подгонять результаты своих трудов под эти критерии, чтобы улучшить свои показатели, но это не улучшит качество самих исследований.

Существует большое количество более подробных публикаций на тему проблем в базах данных цитирования. Даже беглого взгляда на работу Фридмана Маттерна (Friedmann Mattern) достаточно, чтобы перестать слепо доверять автоматически рассчитанным показателям из баз данных.

Мы можем сделать общий вывод: оценивание исследований, основанное на каком-то наборе данных, не может быть качественное самих этих данных. К сожалению, с момента появления подобных баз эта ситуация не становится лучше, и, кажется, нет ни одной организации, которая занималась бы решением этой проблемы. Одним примером, достойным упоминания, является сайт DBPL, поддерживаемый Михаэлем Лейем (Michael Ley) из Университета города Триер. Этот ресурс в отличие от других позволяет связаться с автором, если возникает необходимость исправить ошибку. Но на этом сайте ведется только перечень публикаций, а не цитирований. Последние усилия цифровой библиотеки ACM, упомянутой ранее, также направлены на быструю реакцию на критику и исправление ошибок.

Критический анализ и внимание к индикаторам способны сделать получение данных более прозрачным.

7. Любая характеристика оценивания, особенно количественная, должна основываться на прозрачных, известных всем критериях.

Эта рекомендация все еще актуальна, если говорить о главных

базах данных. Методики, по которым Google Scholar и ISI выбирают документы и цитаты, до сих пор не опубликованы и являются предметом дискуссий. Люди, принимающие решения относительно эффективности того или иного исследования, должны учитывать это.

Шаблоны публикаций разнятся в зависимости от дисциплины, поэтому вновь скажем о том, что нельзя судить об одних вещах по правилам, принятым для других.

8. Сравнивать между собой различные дисциплины, используя количественные показатели, нельзя.

Это правило можно применить также и к принципам оценки деятельности исследовательских лабораторий и институтов, а не только отдельных ученых.

V. ИСПОЛЬЗОВАНИЕ ISI

Одна из проблем, заботящих CS-ученых, заключается в тенденции использовать для оценивания базы данных, которые не покрывают адекватно сферу CS, например, базу данных Thomson Scientific's ISI Web of Science. Этот ресурс был разработан с прицелом на естественнонаучные дисциплины, и если их представители в целом довольны этой базой данных, то для оценки исследований в CS она не подходит.

Конкретно проблема выражается в том, что именно считается ISI заслуживающим подсчета — критерии отбора изданий произвольны и непрозрачны. В списке нет многих CS-конференций и большинства книг, в то время как туда без разбора включены другие публикации. В результате ученых в области CS опускают «ниже плинтуса». Никлаус Вирт, лауреат премии Тьюринга (высшая награда в CS, эквивалент Нобелевской премии в других дисциплинах) и автор языка Паскаль, упоминается в базе данных ISI лишь как автор нескольких незначительных статей — ни слова о его историческом «Описании языка Паскаль» (1970 г. в соавторстве с Катлин Йенсен, Kathleen Jensen), изданном в виде книги и ставшем одной из самых известных публикаций в CS. Также в этой базе данных не фигурирует эпохальная се-

рия книг Дональда Кнута (также лауреата премии Тьюринга) «Искусство программирования», которую цитируют более 15 тысяч источников базы Google Scholar и которая по опросам различных ученых в области CS считается самой важной публикацией в их науке. Нет там и трех его наиболее цитируемых по версии Google статей.

Свидетельствует о небезгрешности ISI в оценке публикаций по CS и низкий уровень «внутреннего цитирования» — процент ссылок друг на друга публикаций определенной тематики. Если внутреннее цитирование для физики или химии в базе данных ISI превышает 80%, то для CS составляет всего лишь 38%.

Еще один пример — серия *Lecture Notes in Computer Science* (LNCS), выпускаемая издательством Springer. До 2006 года база данных ISI классифицировала эту выдающуюся серию как рядовой журнал! LNCS — это колossalный ресурс, который предлагает оперативные публикации трудов и докладов конференций. Абсурдно смешивать все это многообразие работ в одну кучу, именуя просто журналом, особенно с тех пор, как ISI вообще игнорирует отдельные топ-конференции, если их труды не печатаются в серии LNCS:

- Международная конференция по Программной инженерии (ISCE), главная конференция в области, включенной в список ISI, никак в этой базе данных не отражается;

- Зато в ISI индексируются публикуемые в специальной серии LNCS работы, не прошедшие отбор на ISCE и обсуждаемые на отдельной секции конференции как сырье идеи.

Другой пример. ISI индексирует такое издание, как *SIGPLAN Notices*, издаваемое Programming Languages group ассоциации ACM. Это нереферируемое издание, в обычных выпусках которого публикуются заметки и письма, а в специальных выпусках — труды таких конференций, как POPL и PLDI. Представляете, статьи из POPL считаются в ISI эквивалентными по важности тем письмам читателей, которые опубликованы в обычных выпусках SIGPLAN Notices!

Составители базы данных плохо разбираются в предмете Computer Science. В списке 50 наиболее цитируемых работ по теме CS включена статья «Хемометрика в продовольствии» из журнала «Хемометрика и умные лабораторные системы», хотя само название статьи, не говоря уже о ее содержании, не имеет никакого отношения к CS. И это не единичный случай. Некоторые другие работы из этого списка, пусть и относятся к CS (обычно к каким-то специализированным областям, а не к основным направлениям этой науки), но, по мнению самих CS-исследователей, не являются выдающимися. Ужасает сравнение базы данных ISI с базой данной CiteSeer в разделе CS. В «Список наиболее цитируемых работ» (этот список также заслуживает определенной критики) последней включено множество публикаций, известных всем компьютерным специалистам, — этот список не имеет ни одного пересечения с таким же списком ISI!

В статье Мерлете и др. (Merlet [et.al.]) отмечается, что журнал, находящийся на первом месте в рейтинге ISI, занимает только 195-ю позицию в рейтинге CiteSeer, в то время как лучший по мнению CiteSeer журнал по CS, занимает, соответственно, 26 место в ISI. С одной стороны, это может быть доводом в пользу того, что обоим рейтингам не стоит доверять, анализ различий позволяет сделать вывод, что это просто отражение полного несоответствия базы данных ISI с истинным положением дел в CS.

Список наиболее цитируемых ученых, составленный ISI, свидетельствует о полном игнорировании составителями базы данных реалий этой науки. Да, там есть немало имен великолепных ученых в области компьютерных технологий, но нет упоминания о таких светилах как Вирт, Парнас, Кнут и всей десятке победителей премии Тьюринга за 2000–2006 годы за исключением одного. Это Рональд Ривест (Ronald Rivest), первая буква его фамилии — R фигурирует в названии криптографического алгоритма RSA, между тем, там нет Ади Шамира (Adi Shamir), — буква S в

аббревиатуре RSA — еще одного уважаемого ученого в области Теории CS.

Можно предположить, что в той части CS, которая изучает Теорию (эта часть ближе к давно устоявшейся математике, чем часть, изучающая Системы), применение базы данных ISI может дать лучшие результаты. Однако разница с математикой все равно остается очень большой из-за ошибок составления базы данных, находящихся в ее основе.

Так как механизмы ISI не предоставляют прозрачных способов для оценки сообществом и исправления ошибок, не похоже, чтобы ситуация могла поменяться. К сожалению, выбор альтернативы усложняется дефицитом подобных проектов. Google Scholar не борется с ссылками авторов на свои же работы (в CiteSeer этот механизм предусмотрен), ни одна из этих баз данных не публикует четких критериев, по которым она включает публикации в свой список. И все же все эти недостатки не столь существенны по сравнению с проблемами при использовании ISI Web of Science:

9. В оценке публикаций и цитирований по большинству разделов Computer Science база данных ISI Web of Science является неадекватной и не должна быть использована. В число альтернатив ей необходимо внести Google Scholar, CiteSeer и (потенциально) Электронную библиотеку ACM.

Все, кто причастен к оцениванию, должны понимать, что попытки использовать ISI в целях оценивания деятельности в сфере CS вызовут массовое несогласие и могут спровоцировать неприятие всех количественных критериев оценки, включая даже более разумные. Конечно, подобная реакция является чрезмерной, однако, причина ее понятна людям, принимающим решение об оценивании дисциплин. Стоит также сказать, что ученым из других областей науки не стоит навязывать ученым в CS методы оценки исследований, совершенно для этого не подходящие.

Однако даже определив все недостатки базы данных ISI, стоит от-

метить, что систематические исследования различных баз данных цитирований показывают, как сильно различаются в них результаты оценки ученых, в частности, ученых в CS. Поэтому часть авторов самостоятельно выбирает для себя базу данных, на которую стоит ориентироваться, и работает исключительно с ней. Однако такой подход не универсален и, возможно, когда-нибудь появится программный продукт, который позволит применять его автоматически. Пока же стоит помнить о недостатках исходных данных популярных баз и стараться не принимать судьбоносных решений, опирируя только результатами из них.

VI. ФОРМУЛА ОЦЕНИВАНИЯ

В последнее время принято использовать числовые значения для оценки известности (влияния) статьи, основываясь на значениях из баз данных. Особенно часто говорят об h -индексе (вычисляется, как самое большое число n , при котором $C(n) \geq n$, где $C(n)$ — счетчик цитат для n -ной по рейтингу публикации автора). В пользу применения h -индекса говорит то, что его значения довольно хорошо коррелируют с другими показателями успешности ученого (например, получением Нобелевской премии). Однако, насколько нам известно, такая связь с исследованиями в CS не была изучена. Существуют такие вариации:

■ индивидуальный h -индекс, который высчитывается делением h -индекса на количество авторов и лучше отражает личный вклад;

■ g -индекс, равный самому большому n , при котором n верхних по рейтингу публикаций имеют (все вместе) как минимум n^2 цитирований. Это устраняет еще один недостаток h -индекса — невозможность распознать чрезвычайно важные публикации (если на вашу вторую по цитируемости работу ссылается 100 источников, h -индекс никак не учитывает, сколько ссылалось на первую — 101 или 15 000).

Интернет ресурс «Publish or Perish» («Публикуй или погибнешь», www.harzing.com/resources.htm#/pop.htm) подсчитывает эти ин-

дексы по данным Google Scholar. Но такие индексы по определению достоверны ровно настолько, насколько достоверны сведения в базах данных, по которым их вычисляют. Поэтому полученные таким путем результаты необходимо вручную проверять, чтобы верно определить контекст и избежать искажений. Контрпродуктивно и полностью отказываться от практики использовать такие подсчеты, и слепо ими пользоваться. Приходится признать, что ничто пока не может заменить аккуратное использование таких формул, дополненных другими источниками, такими как оценки коллег по специальности.

VII. ОЦЕНИВАНИЕ ОЦЕНОК

Зачастую, когда ученый негативно отзывается о каком-либо способе оценивания, его подозревают в старомодности и нежелании двигаться в ногу со временем или видят в нем неудачника, который не в силах смириться с отрицательной оценкой его деятельности. Чаще всего всё совсем не так. Как мы уже отмечали в начале этого материала, любой ученый понимает, что оценивание — это неотъемлемая часть научной жизни, а неприятие в нем вызывает не оценивание как таковое, а некорректные подходы, часто применяющиеся в этом процессе.

Человека, занимающегося наукой, жестоко муштруют. Ты обязан всесторонне исследовать любую гипотезу, неоднократно повторить любой эксперимент, отыскать независимое доказательство любой теоремы. Естественно, ученые предполагают, что процессы, влияющие на их деятельность в научном учреждении, разрабатываются по аналогичным стандартам и тщательно прорабатываются. И так же, как в споре с аспирантом они не могут позволить себе навязывать некорректное с точки зрения науки мнение, аргументируя его лишь своим старшинством, так и в подходе к оцениванию своей работы они не позволят университетскому начальству навязывать некорректный механизм оценки, основанный только на воле руководства. Развитие современной науки требовало соблю-

дения принципа коллективной самооценки, который следует применять и дальше, несмотря на развитие новых способов оценивания. В первую очередь, следует убедиться, что использующиеся методы оценки основываются на открытых и понятных критериях. Кроме того, они должны соответствовать действительному порядку вещей и постоянно анализироваться. Это тем более справедливо для CS — науки, которая стремительно развивается.

10. Принятые критерии оценивания сами должны быть подвержены внимательной рецензии и оцениванию.

Для того чтобы этот процесс завершился успешно, необходимо искреннее стремление ученых к открытости и самосовершенствованию.

Этот обзор представляет наши самые общие выводы. Необходимо пристально рассмотреть, изучить негативные реакции на предложенные нами новые техники оценивания. Они не являются отрицанием оценивания как такого, но призывают к более взвешенному, профессиональному подходу. Как принято шутливо говорить, есть две новости, плохая и хорошая. Плохая новость: нет простой формулы, нет пресловутой серебряной пули — инструмента, который мог бы выставить магический балл, справедливо характеризующий ученого. Хорошая новость: в нашем распоряжении как никогда много инструментов, сочетание которых может помочь сформировать правдивую картину эффективности исследования в области CS. И мы обязаны использовать эти инструменты так же внимательно, как мы выполняем свою работу ученых.

**Бертран МЕЙЕР,
Кристина ЧОППИ,
Ян Ван ЛИВЕН,
Йорген СТАУНСТРУП**