

ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата физико-математических наук
Перпера Евгения Михайловича
на тему: «Поиск подслова в множестве слов и его приложение
к семантическому анализу текстов»
по специальности 01.01.09 - «дискретная математика и
математическая кибернетика»

Основная часть диссертации Перпера Евгения Михайловича является исследованием в области теории управляющих систем и теории баз данных. Решается задача поиска подслова в наборе слов. Предполагается, что все слова имеют одинаковую длину n , а буквы представляют собой натуральные числа. Если слово-запрос имеет длину меньше n , то к нему приписываются нули так, чтобы преобразованное слово имело длину, равную n . Евгений Михайлович решает эту задачу в нескольких различных постановках и ищет оптимальные по памяти и сложности алгоритмы. Естественно оптимальное решение этой задачи представляет не только научный, но и практический интерес, так как такого рода задачи встречаются в реальной жизни и до сих пор актуальны. Автор избрал в качестве базы для решения этой задачи информационно-графовый подход.

Информационно-графовая модель зарекомендовала себя как очень удобный способ получения верхних и нижних оценок в задачах информационного поиска. Об этом свидетельствуют многочисленные работы Гасанова Эльяра Эльдаровича и его учеников.

Последняя глава диссертации посвящена семантическому анализу текста на русском языке. При этом в качестве текстов выбраны нормативные акты. То есть текст описывает некоторый алгоритм принятия решения.

Подход к решению этой задачи основывается в том числе и на методе моделирования логических процессов, предложенный Александром Сергеевичем Подколзиным. Актуальность этой тематики подтверждает тот факт, что до сих пор нет нормального автоматического инструмента для анализа законов, и многие команды вручную разрабатывают так называемые ERP-системы (Enterprise Resource Planning - управление ресурсами предприятия). Евгения Михайлович в диссертации показывает, как можно сделать эту процедуру автоматической.

Диссертация состоит из введения, четырех глав, двух приложений и заключения.

Во введении формулируется постановка задачи, обосновывается ее актуальность, и приводятся результаты. Также сделан хороший обзор литературы по данной теме.

В первой главе рассматривается задача поиска слов из заданного множества, которые содержат подслово-запрос. Вводятся ограничения на использование различных вычислительных функций, которыми можно пользоваться для решения поставленной задачи. В предложенных ограничениях, удалось получить алгоритмы минимальной сложности. Также были получены нелинейные нижние и верхние оценки на память, необходимые минимальному по сложности алгоритму. Евгений Михайлович нашел порядок роста памяти равный pn^2 , где n это длина слов, а p их количество.

Во второй главе решается чуть более сложная задача. В качестве ответа нужно выдать не просто набор слов, в котором встречается подслово-запрос, но и позицию (номер буквы) с которого это подслово начинается. Удалось найти линейную нижнюю оценку на память минимального по сложности алгоритма, решающего эту задачу, при использовании другого набора вычислительных функций. Кроме этого приводится алгоритм, который не более чем в 3 раза хуже оптимального по сложности, но при этом имеет линейную оценку на используемую память (pn).

В третьей главе постановка задачи совпадает с постановкой задачи первой главы, за исключением одного ограничения. Длина подслова-запроса известна и передается вместе с ним. Это усложняет нахождение алгоритма, так как запрещается смотреть больше букв, чем есть в подслове. Можно считать, что если алгоритм спросит букву, позиция которой больше длины подслова, то случится ошибка обращения к памяти. Таким образом, алгоритм, решающий задачу, должен учитывать длину слова, что естественно накладывает определенные сложности. Евгений Михайлович предложил серию различных алгоритмов, зависящих от объема памяти и сложности. Можно отметить результат, полученный в теореме 6, а именно что если объем памяти не меньше $\frac{(1+\epsilon)pn^3k}{6}$, то верхняя и нижняя оценка сложности отличаются всего на 1, где ϵ это произвольное положительное число, а k количество букв в алфавите.

Четвертая глава посвящена построению модели юридического документа по тексту. Нужно перевести текст закона на язык логики предикатов. Процедура перевода происходит в 4 этапа: синтаксический анализ текста; отождествление сущностей; построение формулы логики предикатов; построение модели вычисления объекта. Приемы синтаксического анализа описаны в приложении 1, а в приложении 2 приведен пример их использования. Считаю, что автор привнес существенный вклад в автоматизацию семантического анализа документов.

Основные результаты диссертации являются новыми, сопровождаются строгими математическими доказательствами. Результаты получены автором самостоятельно. Они опубликованы в научных журналах, определенных Положением о присуждении ученых степеней в МГУ имени М.В. Ломоносова, а также неоднократно докладывались на конференциях. Автореферат диссертации в полной мере отражает ее содержание. Считаю, что диссертация является законченным научным исследованием и представляет как теоретическую, так и практическую ценность.

Вместе с тем, в работе имеется ряд недостатков.

Прежде всего, отмечу, что в тексте диссертации не хватает вспомогательных предложений, которые бы подчеркивали результаты. Читателю приходится самому делать выводы из теорем, хотя автор мог бы помочь читателю это сделать, написав одно два предложения, поясняющее результат. Для примера рассмотрим формулировку теоремы 6. В этой теореме получены два результата оценивающие сложность решения задачи в зависимости от ограничений на объем. В этой теореме есть очень интересный результат о верхней и нижней оценке на сложность, который нужно додумать читателю самому. Верхняя и нижняя оценка отличается всего на константу 1, что случается достаточно редко, и стоило бы отдельно подчеркнуть этот факт. Так же хочу отметить, что алгоритмам не хватает иллюстраций. Можно было схематично пояснить алгоритм в виде картинки, к чему автор неоднократно прибегает во время дискуссии, но, к сожалению, не внес эти схемы в саму диссертацию. Естественно из-за этого усложняется понимание идеи доказательства теоремы. Выбранные обозначения не говорящие. Читателю тяжело понять, чем отличается отношение ρ_1 от ρ_2 и приходится возвращаться к определениям. Можно было использовать другие обозначения, в том числе и многобуквенные. Это замечание направлено на доработку обозначений в дальнейших работах автора.

Не обошли стороной и опечатки в тексте диссертации. Перечислю пару примеров таковых:

1. На странице 12 в самом конце абзаца, предшествующего формулировке теоремы 4, вместо $R(I_2, F_2, x)$ упомянуто $R(I_2, F_1, x)$.
2. На странице 56 в последнем предложении третьего абзаца, считая снизу, должно быть ρ_3 , а не ρ_1 .
3. В формулировках леммы 12 на странице 57 и леммы 13 на странице 61 вместо множества X' упоминается множество X .

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломо-

носова к работам подобного рода. Содержание диссертации соответствует паспорту специальности 01.01.09 - «дискретная математика и математическая кибернетика» (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова.

Оформление диссертации соответствует приложениям № 5, 6 Положения о диссертационном совете Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Перпер Евгений Михайлович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.09 - «дискретная математика и математическая кибернетика».

Официальный оппонент:

кандидат физико-математических наук

Плетнев Александр Андреевич

Контактные данные:

тел.: 7(967)1769263, e-mail: aa.pletnev@yandex.ru

Специальность, по которой официальным оппонентом защищена диссертация:

01.01.09 - «дискретная математика и математическая кибернетика»

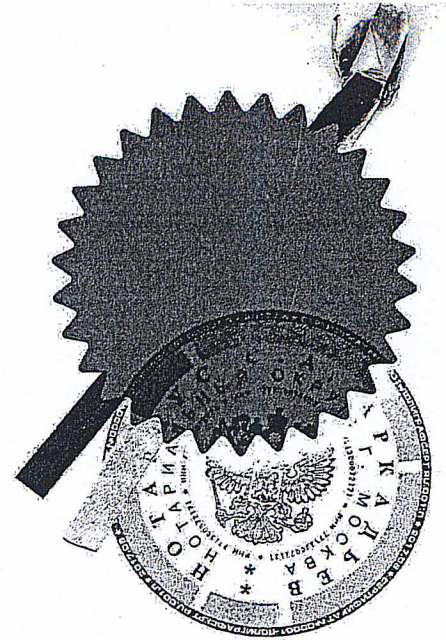
Адрес места работы:

109052, г. Москва, ул. Нижегородская, д. 66, кв. 16,

ИП Плетнев Александр Андреевич

Тел.: 7(967)1769263, e-mail: aa.pletnev@yandex.ru

Подпись А.А. Плетнева удостоверяю:

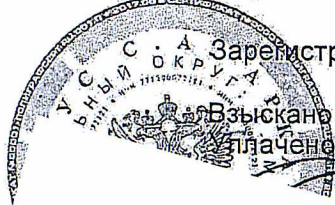


Российская Федерация

Город Москва

Второго апреля две тысячи восемнадцатого года

Я, Аркадьев Сергей Александрович, нотариус города Москвы, свидетельствую подлинность подписи Плетнева Александра Андреевича. 14 октября 1989 года рождения, паспорт гражданина Российской Федерации 45 10 598625, выдан отделением по району Нижегородский ОУФМС России по гор. Москве в ЮВАО 13 ноября 2009 года, код подразделения 770-109. Подпись сделана в моем присутствии. Личность подписавшего документ установлена.



Зарегистрировано в реестре: № 77/32-н/77-2018-1-626.

Взыскано по тарифу:

100 руб. 00 коп.

Плачено за оказание услуг правового и технического характера: 1000 руб. 00 коп.

С.А. Аркадьев

[Handwritten signature]

[Handwritten signature]