ЛИНГВИСТИКА ТЕКСТА

АВТОМАТИЧЕСКИЙ ОТБОР РЕЛЕВАНТНОЙ ИНФОРМАЦИИ ИЗ ИНФОРМАЦИОННОГО МАССИВА ПАТЕНТНЫХ ТЕКСТОВ[1]

О.И. Бабина

1. Введение

Поиск и анализ патентных формул является неотъемлемой частью патентных исследований. Патентная экспертиза осуществляется в несколько этапов, основными из которых являются поиск и отбор патентов, имеющих отношение к объекту, подлежащему проверке, с последующим их анализом.

Исследования осуществляются в трех основных аспектах: техническом, правовом и экономическом. С учетом единства этих аспектов, различают три основных группы целей патентного поиска[2]:

Установление *условий реализации прав патентообладателей*;

Установление *объема прав патентообладателя;*

Установление *уровня технических решений*, среди которых в свою очередь выделяют:

Прогнозирование развития в пределах предметной области;

Проверка на патентоспособность изобретения;

Проверка на патентную чистоту изобретения.

Мы сосредоточимся на последней группе целей. Часто поиск в этом случае проводится с целью отыскания охранных документов, под действие которых может подпадать вновь заявленное техническое решение. При поиске необходимо, просмотрев все действующие патенты, выявить документы, в которых описываются признаки изобретения, совпадающие с признаки проверяемого объекта.

Традиционно поиск осуществляется по предметным рубрикам международной или национальных классификаций изобретений. Дальнейшему анализу подвергаются формулы изобретений патентов; анализ осуществляется экспертами вручную.

Мы предлагаем способ отбора патентов на этапе поиска, использующий методы лингвистического анализа текста при индексировании информационного массива, что позволяет проводить более «тонкий» отбор при поиске с целью установления уровня технических решений.

Разработка и проверка метода отбора патентных формул осуществлялся на материале текстов патентов на способ в предметной области фармакология.

2. Формирование поисковых образов документов в информационном массиве патентных формул

Векторное индексирование, с последующим вероятностным определением степени релевантности документов поисковому запросу, является наиболее распространенной в настоящее время моделью представления поискового массива. Результатом выполнения запросов к такой системе является ранжированный список текстов предположительно релевантных запросу документов. Применение лингвистического компонента в этих системах, чаще всего, сводится к *морфологическому* анализу отдельных слов и их лемматизованному представлению в поисковых образах документов (ПОД). Однако в последнее время исследователи обращают все больше внимания на то, что при проектировании систем поиска можно учитывать и более глубокие уровни языка, увеличивая таким образом показатели эффективности этой системы. В структуру многих систем сегодня включены *синтаксические* анализаторы. Они служат, в одних случаях, для выявления более сложных, чем отдельные словоформы, именных групп (хотя детальность такого синтаксического разбора обычно незначительна). Как правило, разбор ограничивается выявлением границ именной группы (ИГ) по методу сопоставления с эталоном, используя частеречные шаблоны[3], или распространяется далее до выделения ядерного элемента-существительного, образующего эту группу[4]. В других случаях, делается более глубокий анализ, и выделяются связи между лексическими элементами внутри ИГ[5]. В последних моделях появляется возможность не только проводить поиск отдельных слов или словосочетаний из словаря, но также учитывать степень сходства словоформ по зависимостям между ними. Это позволяет отойти от предположения о несвязности терминов и увеличивает точность поиска системы. В ряде исследований делается попытка охватить также и *семантические* особенности текста. Так, существуют модели, в которых ПОД строится посредством выделения в ходе синтаксического анализа ИГ (или их элементов) с последующим наложением на концепты метатезауруса. В результате поисковый образ формируется не с помощью списка лексических элементов, а на основе концептуального представления текста документа.

Однако в подавляющем большинстве таких моделей поиска лингвистический инструментарий применяется исключительно для выявления и представления ИГ. При этом элементы предложения, на которых лежит основная смысловая нагрузка при описании внеязыковой ситуации, – предикаты – остаются без внимания. Мы полагаем, что для поиска документов и дальнейшего извлечения информации из них необходимо учитывать также и информацию, заложенную в предикатах.

Для индексирования документов мы используем процедуру анализа текста, на выходе которой получается предикатная структура документа. В дальнейшем эта предикатная структура будет использоваться в качестве ПОД документа.

За основу индексирования мы берем процедуру анализа, составленную ранее для анализа патентов на устройства[6]. Основой для выполнения автоматического анализа является лексикон, где каждый предикат описывается с помощью его семантического класса, набора семантических ролей, набора линейных шаблонов, представляющих способы линейной реализации семантических ролей в тексте формулы изобретения, и способов заполнения семантических ролей в терминах синтаксических структур[7]. Предикаты в патентах на способы специфичны по своим характеристикам и поведению[8], поэтому подготовительная работа по описанию предикатов в лексиконе является необходимым этапом успешного анализа.

Анализ включает ряд последовательно выполняющихся этапов:

а) разметка текста: выполняется грубое деление текста на блоки на основе информации о пунктуации, визуальной организации текста и т.д.;

б) морфологический анализ: текст размечается с помощью набора супертэгов, приписанных каждому слову в словаре. Супертэги помимо информации о части речи включают информацию о семантическом классе слова. После разметки выполняется этап разрешения неоднозначности, так как ряду слов может быть приписано несколько меток. Правила разрешения неоднозначности зависимы от объекта изобретения, поэтому правила для патентов на способ подверглись значительным изменениям;

в) синтаксический анализ: в тексте выделяются синтаксические блоки, соответствующие простым именным группам, сложным именным группам, наречным оборотам, инфинитивным оборотам, герундиальным оборотам. Структура синтаксических блоков мало отличается для различных подъязыков в пределах одного естественного языка. Поэтому большая часть правил, используемых при анализе текстов патентов на устройства, оказалась пригодной для анализа патентов на способы. Были добавлены лишь отдельные правила-шаблоны для распознавания именных групп[9];

г) семантический анализ: включает выделение предикатов в тексте, распознавание границ предикатных конструкций и соотнесение каждого синтаксического блока в пределах одной предикатной конструкции с соответствующей ему семантической ролью. Алгоритм анализа на этом этапе использует описания предикатов из лексикона.

Правила анализа на каждом этапе описываются в формализме ЕСЛИ-ТО-ИНАЧЕ. В результате анализа текст формулы изобретения представляется в виде графа связанных между собой предикатно-аргументных структур, где в качестве узлов выступают предикаты и их актанты, а отношения между узлами соответствуют семантическим ролям, которые актанты выполняют.

Полученное в ходе автоматического анализа представление используется в системах автоматического перевода, синтеза текста патентных формул. Однако процедура была расширена с целью учета в индексе документов макроструктуры патентной формулы.

В организации формулы изобретения патента на способ можно различить следующие разделы:

PURPOSE: заголовок способа, где описывается цель способа (лечение заболевание, получение нового вещества и т.д.);

IMPLEMENTATION: список компонентов способа, где раскрываются действия направленные на реализацию описываемого способа.

Эти два раздела формально разделяются предикатом comprising (или выражением comprising the step(s) (of)). Второй раздел, в свою очередь, можно условно разделить на зоны, каждая из которых соответствует одной операции способа. Центральным и первым элементом в каждой зоне является предикат, обозначающий действие-компонент способа, выраженный герундием или существительным. Перечисленные особенности дают формальные признаки для выделения указанных разделов и зон.

При проведении индексирования на этапе выделения предикатов основному предикату в каждом разделе приписывается метка: а) PURP (предикат, описывающий цель способа); или б) IMPL (у), где у – номер зоны (компонента способа) по порядку (предикат, обозначающий компонент способа).

Для формирования ПОД документа представляется необходимым установление кореферентных связей между элементами способа (названиями веществ, заболеваний и т.д.), так как в ходе патентной экспертизы информация о том, выполняется ли действие с одним и тем же или разными элементами релевантна. Поэтому в процедуру анализа добавлен блок автоматической расстановки кореферентных связей.

Объекты-участники способа выражаются в тексте с помощью именных групп, поэтому только эти синтаксические структуры рассматриваются на предмет расстановки кореферентных связей. Для обозначения кореферентных ИГ будем помечать соответствующие группы одинаковой цифрой.

В ряде случаев задача расстановки кореференции тривиальна, так как некоторые актанты, заполняющие в предикатных структурах на выходе процедуры анализа семантические валентности двух предикатов, на самом деле в линейном тексте формулы изобретения используются лишь однажды. А так как каждый элемент текста в ходе морфологической разметки помечается супертэгом с уникальным номером, кореференция в таких случаях расстанавливается автоматически: все актанты, помеченные одинаковыми номерами, кореферентны.

Задача, которая не так тривиальна, сводится к определению кореферентных связей между именными группами, встречающимися в тексте несколько раз. Случаи, в которых необходимо восстановить кореференцию, можно разбить на две категории:

а) для ссылки на ранее упомянутый объект используется ИГ, где главное существительное совпадает с главным существительным в ранее использованной ИГ;

б) для ссылки на ранее упомянутый объект используется местоимения.

Так как в силу особенностей подъязыка патентных формул местоимения обладают крайне низкой частотностью, мы рассматриваем лишь первый случай. Восстановление кореференции осуществляется с помощью составленных нами правил эвристики. Именные группы ИГ0, которые потенциально кореферентны одной из ранее упомянутых в тексте формулы ИГ1, рассматриваются в линейном тексте формулы слева направо. Потенциально кореферентной одной из ранее упомянутых ИГ считается ИГ, начинающаяся с определенного артикля или слова 'said', соответствующего в нашем корпусе определенному артиклю. Поиск кореферентной связи осуществляется по линейной текстовой последовательности, которая предшествует рассматриваемой ИГ. Причем поиск кореферентной ИГ1 осуществляется поблочно: вначале в пределах того же блока слева направо, затем в блоке, предшествующем данный, слева направо и так далее, пока не будет рассмотрен начальный блок пункта формулы (раздел PURPOSE). Порядок правил в эвристике важен и начинается с правила, включающего наиболее строгое условие (полное совпадение). В последующих правилах условия совпадения ИГ постепенно ослабляются.

Ключевым понятием в эвристике является понятие головы ИГ. Определение головы задается отдельным набором правил, составленных на основе анализа ИГ в корпусе текстов и представленных, как и остальные правила анализа, в формализме ЕСЛИ-ТО-ИНАЧЕ.

Результат применения данной процедуры анализа можно схематически представить на примере. Шаг способа, представленный в формулы изобретения текстовой последовательностью «*mixing a pharmaceutically acceptable surfactant with with said nanoparticles in said pharmaceutically acceptable aqueous solvent*», после прохождения через процедуру анализа будет представлен следующей предикатной структурой (головы именных групп подчеркнуты и помечены индексами, используемыми для установления кореферентных связей):

[mixing]:
SEM-CLASS: Interaction
TYPE: IMPLEMENTATION (2)
CASE-ROLES:
1 [a pharmaceutically acceptable surfactant$_1$] // <direct-obj>
2 (with) [said nanoparticles$_2$] // < indirect-obj>
3 (in) [said pharmaceutically acceptable aqueous solvent$_3$] //< place>

Автоматическому анализу по представленной схеме подвергаются единовременно все документы информационного массива. Полученные в результате анализа представления патентных текстов в виде набора предикатных структур используются в качестве ПОД документов при последующем поиске.

3. Формирование поискового образа запроса к информационному массиву патентных формул

Для извлечения необходимой информации из массива документов необходимо представить информационные нужды пользователя в формате, доступном для «понимания» системы. Определение нужд пользователя осуществляется в ходе формирования поискового образа запроса (ПОЗ). В современных поисковых системах для формулировки запроса используется ряд подходов.

Формулировка в виде набора ключевых слов, соединенных логическими операторами: наиболее часто применяющийся подход, используемый в системах с координатным индексированием. При обработке такого запроса отыскиваются документы, для которых логическое выражение в ПОЗ является истинным. В усложненном варианте каждому термину запроса назначаются веса, и ранжирование найденных документов осуществляется на основе приписанных каждому термину весов;

Формулировка в виде предложения на естественном языке: при обработке таких запросов могут использоваться те же методы обработки, что и в предыдущем подходе: из запроса выделяются именные группы, которые в дальнейшем играют роль ключевых слов. Этим словам, как правило, приписывается вес. Причем расчет весов может осуществляться различными способами: по тому, сколько раз термин встречается в запросе[10], на основе принадлежности термина полям-категориям, заданным в тезаурусе и имеющим разные коэффициенты веса[11] и т.д. Использование запросов на ЕЯ дает возможность получить представление о синтаксической связности отдельных ключевых слов запроса, увеличивая точность.

Использование в качестве запроса документ (или набор документов): такие запросы использу-

ются в системах, поддерживающих функцию ассоциативного поиска. Ответом на такой запрос являются источники, «похожие» на документ-запрос и найденные с помощью алгоритмов ассоциативного поиска, которые, как правило, основаны на частотном анализе текстов.

При проведении патентных исследований стоит следующая задача: имея описание нового объекта на естественном языке, проверить, нет ли других запатентованных изобретений, которые могут накладывать ограничения на его использование.

Зачастую способ представления поисковых образов документов в информационном массиве совпадает со способом представления запроса. Это дает возможность сравнивать однородные по составу данные и выявлять степень сходства между структурами, представляющими запрос и документ. В нашем случае способ индексирования, результатом которого является дерево предикатных структур, определяет и форму представления ПОЗ. Мы предлагаем вводить запрос, используя не линейные текстовые вопросы на естественном языке, а последовательно описывая каждую предикатную структуру отдельно. Задание описания изобретения проводится в ходе интерактивного опроса пользователя.

Еще на заре развития систем поиска было отмечено, что характеристики поиска можно улучшить, используя предписывающий язык при формировании запросов. В подавляющем большинстве случаев в качестве такого языка используются информационно-поисковые тезаурусы, где перечислены существительные и именные группы, описывающие объекты, характерные для данной предметной области. Для нашей модели такой инструмент также является полезным. Это обусловлено тем, что с целью расширения прав патентообладателя в тексте патентов используются наиболее широкие определения участников способа. Поэтому в случае, когда в запросе упоминаются более узкие понятия, подпадающие под указанные в патенте, такой патент должен оказаться релевантным запросу. Так патент, в тексте которого присутствует термин «eye disorders» релевантен запросу, в котором указывается «conjunctivitis».

Однако в нашей модели тезаурус, включающий термины предметной области, связанные между собой меронимическими отношениями, не является достаточным словарным инструментарием. Поскольку в патентах на способы основная смысловая нагрузка ложится на предикаты, представляется целесообразным при формировании запроса использовать предписывающий язык, предлагающий пользователю выбор *предикатов* для описания целей и шагов способа. В этом случае есть возможность определить семантический класс и возможных участников описываемой ситуации, используя информацию из словаря предикатов. Пользователю лишь останется явно указать участников ситуации, назначив им соответствующие роли.

Ранее методика, в ходе которой осуществлялось последовательное определение пользователем предикатных конструкций, использовалась для построения структуры описания изобретения с целью автоматического синтеза текста патентной формулы[12]. Эта методика применялась для синтеза формул изобретения на устройства. В патентном праве указано, какие описания должны быть включены в патентную формулу, что и определило наличие следующих этапов ввода: а) ввод названия изобретения; б) его компонентов; в) описание свойств компонентов; г) отношений между компонентами.

Учитывая специфику подъязыка патентов на способы, процедура ввода описания изобретения несколько преобразуется и включает следующие этапы: а) описание цели способа; б) описание шагов способа; в) описание свойств компонентов, участвующих в выполнении шагов способа; г) описание отношений между компонентами. По содержанию шаги (в) и (г) при описании способов идентичны одноименным шагам, выполняемым при описании изобретений на устройство.

В шагах (а) и (б) цели и шаги способа соответственно описываются с помощью предикатных конструкций. Поэтому если в патентах на устройства здесь задавались название устройства и его частей (с помощью именных групп), то в патентах на способы предлагается сначала выбрать предикат. В предлагающийся набор предикатов включены предикаты, которые в корпусе текстов используются для описания целей и шагов способов. Далее для каждого предиката, на основе информации из лексикона, пользователю предлагается фреймовая структура, где в качестве слотов представлен набор возможных валентностей выбранного предиката. Используя тезаурус, пользователю предлагается заполнить слоты, указав участников способа. Фактически, при работе с описаниями способов шаги (а), (б) и (г) становятся идентичными по методу реализации, так как на каждом из шагов проводится описание изобретения посредством задания предикатных структур.

При формировании ПОД мы завели дополнительную характеристику предикатов (TYPE), указывающую на раздел, который данный предикат представляет. Очевидно, что при формировании запроса эта характеристика может приписываться предикатам автоматически, в зависимости от того, на каком шаге осуществляется описание изобретения: шаг (а) соответствует описанию в разделе PURPOSE, шаг (б) – описанию в разделе IMPLEMENTATION. Порядок задания предикатов в ПОЗ приравнивается к порядку осуществления шагов, что дает возможность автоматически нумеровать предикаты в последнем разделе.

Финальным шагом формирования запроса является этап расстановки кореферентных связей, осуществляемых пользователем вручную. Пользователь нумерует существительные, имеющие оди-

наковую морфологическую форму, так, что если два существительных кореферентны, им присваиваются одинаковые номера.

Такое описание приводит к формированию структуры, подобной той, которая получается при индексировании патентных документов. Так, запрос на способ, включающий шаги «*dissolving an amphiphilic heparin derivative in an aqueous solvent*» и «*mixing a surfactant in the solvent*», будет представлен с помощью следующих предикатных структур:

[dissolving]:
SEM-CLASS: Interaction
TYPE: IMPLEMENTATION(1)
CASE-ROLES:
[an amphiphilic heparin derivative]/<direct-obj>
2 (in) [an aqueous $solvent_1$] //<place>
[mixing]:
SEM-CLASS: Interaction
TYPE: IMPLEMENTATION (2)
CASE-ROLES:
1 [a surfactant] //<direct-obj>
2 (in) [the $solvent_1$] //< place>

4. Отбор патентных формул на основе сопоставления ПОЗ и ПОД

Для отбора необходимой информации необходимо сопоставить ПОЗ и ПОД. В нашей модели поисковые образы представляют собой набор предикатных конструкций. Фактически, задача состоит в том, чтобы, имея эталонную предикатную конструкцию, найти в базе индексов патентов «похожую». Традиционно, для определения степени сходства используются неким образом заданные (индивидуально для каждой модели) коэффициенты, называемые *критерием выдачи*. Для определения критерия выдачи необходимо знать, какие характеристики релевантны для отбора. В нашем случае нужно учитывать, что поиск на корпусе патентных формул осуществляется с целью проведения патентной экспертизы.

Исследовав особенности рассматриваемой предметной области и объекта изобретения, мы выделили ряд особенностей, которые представляются важными при сопоставлении предикатных структур.

На уровне предикатной конструкции:

иерархичность семантики предикатов (Pred);

структурное сходство именных групп, используемых для описания участников (NP);

приоритет головного существительного в именной группе (Head);

меронимические отношения между объектами предметной области (Term);

семантическая роль термина, используемого в предикатной конструкции (SemR).

На уровне ПОЗ в целом:

последовательность реализации шагов способа (Seq);

кореференция терминов (Coref).

Степень сходства оценивается по приведенным критериям. Мы рассматривали лишь критерии первого уровня, так как часто сходство по отдельным предикатным конструкциям достаточно для проведения отбора. Таким образом, в качестве поисковой единицы принимается одна предикатная конструкция.

В общем виде процедура сравнения имеет итеративный характер. Каждый шаг представляет собой последовательное рассмотрение одной предикатной конструкции ПОЗ на соответствие ПОД. Существует ряд способов оценить сходство. Наиболее подходящим для нашей модели мы считаем оценку путем подсчета «расстояний» между поисковыми единицами в запросе и индексе документа. Та же идея используется в современных системах машинного перевода, основанного на примерах, на этапе поиска соответствия входного выражения некоторым выражениям из базы примеров. Расстояние в общем виде рассчитывается как сумма взвешенных расстояний по ряду критериев[13]. Для простоты мы считаем все критерии в нашей модели равноценными, поэтому веса не учитываются.

После выполнения этих шагов для каждой предикатной конструкции мы получаем n коэффициентов (n – количество предикатных конструкций в запросе), где каждый указывает на степень сходство по одному из предикатов запроса. Далее необходимо сформировать критерий выдачи, на основании которого и будет проводиться отбор. Коэффициент соответствия для набора предикатов получается аналогичным способом, что и для каждой предикатной конструкции. Коэффициент соответствия представляет собой расстояние между ПОЗ и ПОД, которое определяется суммированием взвешенных расстояний для всех используемых в запросе предикатов. Так как предикаты, обозначающие компоненты способа, при поиске имеют большую ценность, чем те, которые выражают отношения между объектами, веса на втором этапе расчета коэффициента учитываются. Для унификации критерия выдачи, полученного для разных документов, значение коэффициента соответствия нормализуется количеством предикатов в запросе.

Таким образом оценивается каждый документ в информационном массиве. В результате каждый документ соотносится с числовым значением, характеризующим расстояние между документом и рассматриваемым запросом. Далее процедура отбора идентична тем, которые проводятся при традиционном поиске: для отбора документов устанавливается пороговое значение коэффициента и все документы, которые попадают в установленный числовой промежуток, ранжируются в соответствии со значением коэффициента соответствия.

5. Заключение

Мы представили модель отбора информации в узкой предметной области. Преимуществом модели перед традиционно используемыми моделями поиска состоит в том, что делается попытка

учесть не только терминологическое соответствие, но и взаимосвязь между понятиями при поиске. Данная модель рассматривает этап отбора патентов для последующего анализа экспертами. Однако она может быть расширена для проведения автоматизированного этапа анализа технических решений за счет модификации критериев релевантности. Рассмотрение возможностей такого расширения входит в круг ближайших задач.

[1]Работа выполняется при финансовой поддержке Федерального агентства по образованию (грант №А04-1.5-323).

[2] Информационно-поисковые системы и традиционный патентный поиск: Уч. пособие / Под ред. Б.С. Розова. М.: ВНИИПИ, 1987. С. 3.

[3] Lee, Changki, Seungwoo Lee and Gary Geunbae Lee. POSNIR: Probabilistic Natural Language Information Retrieval System. In *Working notes of the Third NTCIR Workshop Meeting Part II: Cross Lingual Information Retrieval Task.* Tokyo, Japan. October 2002. P. 165–171.

[4] Aronson, Alan R., Thomas C. Rindflesch, and Allen C. Browne. Exploiting a Large Thesaurus for Information Retrieval // nls5.nlm.nih.gov/pubs/riao94.pdf; Evans, David A.; Kimberly Ginther-Webster; Mary Hart; Robert G. Lefferts; and Ira A. Monarch. Automatic indexing using selective NLP and first-order thesauri. In *RIAO 91.* P. 624–644.

[5] Matsumara, Atsushi, Atsuhiro Takasu, and Jun Adachi. Structured Index System at NTCIR1: Information Retrieval using Dependency Relationship between Words. In *Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition.* Tokyo, Japan. August 1999. P. 117–122; Smeaton, Alan F., Ruairi O'Donnell and Fergus Kelledy. Indexing Structures Derived from Syntax in TREC-3: System Description. http://citeseer.nj.nec.com/smeaton95indexing.html.

[6] Sheremetyeva S. Natural Language Analysis of Patent Claims. In *Proceedings of the Workshop on Patent Corpus Processing.* Sapporo, Japan. July 2003. P. 66–73.

[7] Sheremetyeva S. A Flexible Approach to Multi-Lingual Knowledge Acquisition for NLG. In *Proceedings of the 7th European Workshop on Natural Language Generation.* Toulouse, France. May 1999. P. 106–115.

[8] Бабина О.И. Грамматические характеристики предикатов формулы изобретения патентов на метод // Вестник ЮУрГУ. Сер. «Лингвистика». 2004. №1. С. 8–12.

[9] Бабина О.И. Специфика процедуры автоматического анализа текстов патентов на метод // Объединенный научный журнал. Москва. Декабрь 2004. №33 (125). С. 62–66.

[10] Fujita, Sumio. Notes on Phrasal Indexing JSCB Evaluation Experiments at NTCIR AD HOC. In *Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition.* Tokyo, Japan. August 1999. P. 101–108.

[11] Youli, Qu, Xu Guowey, and Wang Jun. Rerank Method Based on Individual Thesaurus. In *Proceedings of the Second NTCIR Workshop on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization.* Tokyo, Japan. March 2001. P. 553–558.

[12] Sheremeyteva, Svetlana, Sergei Nirenburg, and Irene Nirenburg. Generating Patent Claims from Interactive Input. In *Proceedings of the 8th. International Workshop on Natural Language Generation.* Herstmonceux, England. 1996. P. 61–70.

[13] Sumita, E. and Iida, H. Experiments and Prospects of Example-based Machine Translation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.* 1991. P. 185–192.