

ПЛАНИРОВАНИЕ ТЕКСТОВ В СИСТЕМАХ ГЕНЕРАЦИИ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

Соколова Е. Г.
РОСНИИ ИИ
sokolova@aha.ru

Болдасов М. В.
РОСНИИ ИИ
boldasov@nm.ru

В статье рассматриваются вопросы планирования текста в системах генерации на естественном языке. Процесс генерации рассматривается как последовательное прохождение обрабатываемыми данными различных информационных состояний в соответствии с этапами генерации. Подробно обсуждаются виды двух информационных состояний – входные данные и планы текста.

Процесс генерации на естественном языке (ГЕЯ) принято представлять с использованием известной в теории информационных систем идеи **конвейера обработки данных**. Обратными связями между явлениями разных уровней в большой степени или полностью пренебрегают. Путем обобщения опыта создания действующих систем ГЕЯ, построена схема, представленная на рис. 1

([Bateman & Zock 2001]), которая отражает общую картину преобразований в системе ГЕЯ. Из схемы видно, что систему генерации удобно реализовывать в виде конвейера, состоящего из трех относительно независимых блоков: планирование содержания текста (макропланирование), планирование предложений (микропланирование) и языковое оформление.

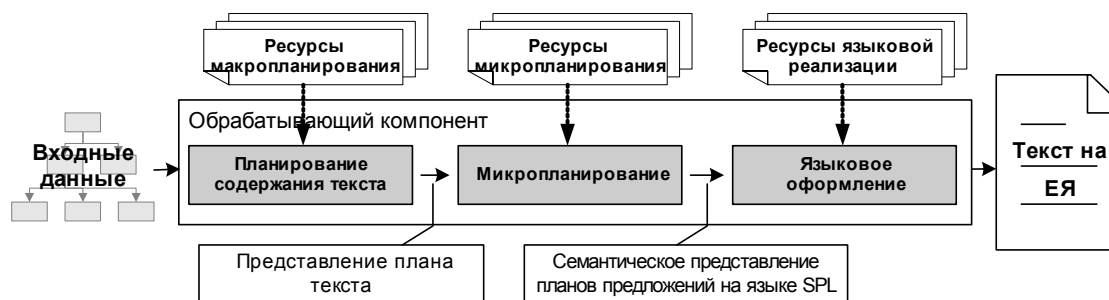


Рис. 1. Схема генерации

Конкретное наполнение этих блоков определяется видом входных и выходных данных, а также видом используемых ресурсов. Процессы заключительного этапа генерации - языкового оформления - достаточно хорошо изучены и для них построены специальные среды их моделирования (наиболее известные из которых – KPML, Serge, RealPro, TagGen). Ресурсы языкового оформления – это описания средств конкретных естественных языков (ЕЯ). Вопросы планирования менее изучены. Ресурсы планирования мало зависят от ЕЯ. Они оперируют с абстрактными понятиями, зависимыми в большей степени от модели предметной области (ПО), в которой применяется генератор.

В соответствии со схемой на рис. 1 мы рассмотрим процесс генерации как последовательное прохождение входными данными следующих информационных состояний:

- Входные данные (начальное информационное состояние)

- Представление плана текста
- Представления планов предложений
- Текст на ЕЯ (конечное информационное состояние)

Дальнейшее изложение имеет следующую структуру. В первом разделе рассматриваются виды входных данных. Во втором разделе мы рассматриваем правила построения планов текста. В заключении резюмируется состояние дел в планировании текстов.

Входные данные

Особенность ГЕЯ состоит в том, что, в отличие от других направлений, например, автоматического анализа или понимания текста, имеющих в качестве входа всегда только текст на ЕЯ, вид входных данных системы ГЕЯ четко не определен. С самой общей теоретической точки зрения считается, что такая система должна принимать на входе четыре составляющих:

- систему знаний, на основе которых будет составляться текст,
- коммуникативную цель этого текста,
- модель адресата текста и
- контекст повествования.

Исходя из гипотезы о наличии такого универсального вида входных данных логично предположить также универсальность и описания процесса генерации в виде конвейера, представленного на рис. 1. В этом случае разнообразие реальных систем должно объясняться тем, что их входные данные представляют разные точки этого конвейера. Однако в реальных системах генерации письменных текстов вход обычно представлен одной составляющей – системой знаний, а параметры коммуникативной цели и адресата либо встроены в систему, либо могут уточняться в процессе генерации. Контекст повествования обычно используется в диалоговых системах для учета результатов предыдущих коммуникаций между пользователем и системой. Опыт разработки реальных систем ГЕЯ показывает, что входные данные в любой системе проходят всю последовательность преобразований, представленную на рис.1. Разнообразие входных данных изменяет лишь объем и характер преобразований, не изменяя их природы. Поскольку процессы, затрагивающие явления различного масштаба и природы, не всегда можно отделить друг от друга, разделение на блоки на рис. 1, в принципе, условно. Иногда эта структура соответствует устройству реальной системы, иногда играет только методическую роль. Вопрос о том, что должна принимать система ГЕЯ на входе, решается в каждой конкретной системе индивидуально. Условно, входные данные можно разделить на две группы: сырые данные и семантические данные.

В сырых данных с точки зрения качественно различных типов информации, которую они могут хранить, можно выделить спектр между двумя типами данных – поток данных и описание постоянных свойств объектов в БД. Сырые данные представляют собой только констатацию наличия определенных объектов, отношений, значений атрибутов, не организованную для передачи адресату в ходе ЕЯ коммуникации.

Поток данных представляет собой обычно численные результаты измерений определенных параметров одного или группы однотипных объектов или явлений через определенные отрезки времени. Объекты и явления могут быть различной природы, например, метеорологические данные (погода), социологические данные (занятость населения), финансовые данные (биржевые сводки), географические данные и др. Потоки данных создаются в процессе автоматического физического мониторинга (физические измерения) или статистического мониторинга (обработка данных, поступающих по электронным каналам) свойств

объекта или явления. Эти данные описывают изменение состояния объекта по заданным параметрам. Отношения между параметрами и объектом или явлением постоянны.

Базы Данных (БД) представляют более общий вид информации об объектах и явлениях. В том случае, если в БД представляются только значения изменяемых параметров, она качественно эквивалентна потоку данных. Но обычно БД описывают как (регулярно) изменяющиеся, так и более или менее постоянные свойства объектов и явлений. Например, БД отдела кадров предприятия состоит из записей сотрудников этого предприятия, включающих дату рождения (постоянный атрибут для отдельного человека), дату приема на работу (постоянный атрибут для данной БД), количество детей (изменяемый атрибут), зарплату (изменяемый атрибут) и т.д. Эти данные описывают объект или явление с разных сторон. Значения атрибутов могут быть связаны с описываемым объектом различными семантическими отношениями, например, член (группы), цвет, количество и т.п.

Базы Знаний (БЗ) не представляет отдельного вида входных данных. БЗ отличаются от БД тем, что имеют особую организацию хранящейся в них информации, позволяющую осуществлять логический вывод, аппарат которого часто также заложен в самой БЗ. В результате, из нее может извлекаться не только информация, эксплицитно заданная, как в БД, но и имплицитная информация, и аппарат вывода может организовывать выбираемые данные согласно заданной коммуникативной цели. В этом случае результаты, полученные из БЗ, нужно рассматривать как семантические данные.

Можно выделить два вида семантических данных – формальные спецификации и семантические представления. Семантические данные отличаются от сырых тем, что они коммуникативно организованы, т.е. подготовлены для передачи адресату в виде некоторой структуры.

Формальные спецификации имеют предметную природу и различные источники. Они формируются:

- при создании программного продукта как побочный результат использования автоматизированных систем проектирования компонентов этого продукта (например, средств проектирования общей архитектуры системы, графических интерфейсов с пользователем, и т.д.);
- в результате вывода информации из БЗ;
- в результате автоматического анализа текстов, например, в ЕЯ-интерфейсе к БД системы InBASE [Жигалов и Соколова 2001], которая автоматически преобразует полученный от пользователя ЕЯ-запрос в запрос на языке SQL. Правильности интерпретации системой пользовательского запроса может контролироваться с помощью системы ГЕЯ,

принимавшей на входе построенный системой SQL-запрос. Она вновь генерирует из него запрос на ЕЯ и предоставляет пользователю для контроля ;

- человеком на специализированных языках, например, SQL, математические и логические выражения, доказательства, написанные на формальном языке и т.д.

Сферой применения формальных спецификаций обычно является человеко-машинная коммуникация.

В отличие от формальных спецификаций **семантические представления** имеют ЕЯ природу. Они ориентированы специально на использование системами ГЕЯ для создания на их основе текстов на ЕЯ. Семантические представления не возникают как побочный выход при выполнении другой задачи. Они создаются вручную, в частности, с использованием специальные пользовательских интерфейсов, в которые имеют заданную модель ПО и предлагают пользователю на каждом шаге построения такого представления инвентарь средств, пригодных для продолжения формирования содержания текста. Например, интерфейс пользователя в системе AGILE [Kruijff et al. 2000], используемый техническими писателями для создания содержания текста инструкции (представлений A-box). Этот процесс опирается на внутреннюю речь человека, как бы проговаривающего про себя будущий текст.

План текста

План текста – это представление информации, составляющей содержание будущего текста, организованное в виде единой структуры, которая может быть отображена в структуру текста на ЕЯ. Структура текста на ЕЯ называется дискурсной структурой, а отношения, образующие эту структуру – дискурсными отношениями. Дискурсные отношения устанавливаются между единицами текста (высказываниями) или фрагментами текста (предложениями, абзацами и т.п.). Они создаются говорящим, преследующим конкретную коммуникативную цель. План текста может отражать как предметную, так и семантическую стороны дискурса. В последнем случае он приближен к структуре текста.

Существует два подхода к построению плана текста: с выделением сообщений (предикативно организованных фрагментов данных, записанных в объектном представлении), и без выделения сообщений. Сообщения в плане текста представляют собой прототипы фраз будущего текста. Независимо от того, имеют ли они предметную природу или ЕЯ семантическую природу, сообщения организуются в структуру текста дискурсными отношениями.

Если план текста строится без выделения сообщений, то структура дискурса представлена иерархией узлов, соединенных, отношениями

атрибуто-объект, меронимии, родовидовыми, логическими, “партиципантными” и т.д. В этом случае план текста напоминает скорее БЗ.

Структура плана текста, построенного без выделения сообщений

Планы текстов без использования сообщений обычно реализуют стратегию изложения **от общего к частному**. Эта структура отражает коммуникативный аспект организации текста – сначала сообщается общая информация, а затем подробности. Такая стратегия изложения характерна для текстов типа отчет на основе данных, включающих взаимосвязанные объекты. По своему характеру отношения между элементами плана текста являются атрибутно-аспектуальными или тезаурусными.

Примером плана текста со взаимосвязанными параметрами типа отчет может служить дерево “топик-коммент” в системе GOSSIP [Иорданская, 1992]. Корень иерархии представляет узел, обозначающий ситуацию, например:

“Использование времени ЭВМ”.

Далее план текста распространяется уточняющими отношениями, исходящими из корня и его дочерних узлов:

*участник (ситуации) :
“пользователи”;*

*аспект (ситуации) :
“нормальная работа ЭВМ”;*

*атрибуто(ы) объекта “пользователи” :
“количество = 2”,
“список = Джесси, Мартин”,
“элемент = Джесси”,
“элемент = Мартин”;*

*действие (Мартина) :
“печать файлов”.*

Действие “печать файлов”, в свою очередь, рассматривается как ситуация, которая может развертываться в зависимые узлы, например:

*аспект (действия “печать файлов”) –
проблемы печати файла, и т.п.*

На основе такого плана может быть построен следующий текст:

The system Uwas used for 7 hours 32 minutes 12 seconds. UThe users Uof the system Uran Ucompilers Uand Ueditors Uduring this time. UThe compilers Uwere run six times, for 47% of the cpu time. The editors were run twelve times, for 53% of the cpu time. Two users, Jessie and UMartiUn, logged on to the UsystemU. UJessie Uused the system for 63% of the time in use. UMartin Uused the system for 40% of the time in use.

Сообщение как единица плана текста

Сообщения – элементарные единицы плана текста. Сообщения объединяют предикативно организованные фрагменты данных, записанные в объектном представлении. Сообщения могут быть как предметной, так и ЕЯ семантической природы.

Сообщения предметной природы создаются на основе входных данных типа поток данных или БД и выражаются на некотором формальном языке, описывающим коммуникативные отношения между *Сущностями*, *Концептами* и *Отношениями* в модели ПО. Использование терминов «сущность», «концепт» и «отношение» имеет здесь скорее интуитивный характер. Под сущностями понимаются категориальные объекты ПО, отражающие объекты реального мира (например, *поезд* или *дом*), место действия и время. Концепты описывают конкретные экземпляры сущности (например, *следующий поезд* или *Дом на Набережной*). Отношения задают связи между Сущностями. Например, *Отношение Время отправления и прибытия поезда* описывает определенный вид связи между поездами и временами.

Приведем пример сообщения, составленного планировщиком системы ГЕЯ на основе таблицы расписания поездов:

```
[message-id: msg02
relation: DEPARTURE
arguments: [departing-entity: CALEDONIAN-EXPRESS
departure-location: ABERDEEN
departure-time: 1000 ] ]
```

Такому плану может соответствовать, например, текст

«Поезд Экспресс-Каледония отправляется из Абердина в 10 часов утра», или текст «Время отправления поезда Экспресс-Каледония – 10.00».

В приведенном примере отношение *Время отправления поезда*, связывает концепты *Поезд Экспресс-Каледония*, *Абердин* и *10 часов утра*.

Как было отмечено выше, в плане текста сообщения представляют собой элементарные единицы, объединяемые дискурсивными отношениями. Например, приведенное выше сообщение может войти в план текста, представленный на рис. 2.



Рис. 2. План текста с использованием сообщений

Из плана на рис. 2 может быть порожден, например, такой текст:

«Из Абердина в Глазго ежедневно ходит 20 поездов. Ближайший поезд – Caledonian Express. Он отправляется из Абердина в 10 утра».

Сообщения в виде семантических выражений представляют собой пропозиции, включающие семантический (лексический) предикат с участниками. Предикат является коммуникативно организованной реализацией предметного отношения, например, для предметного отношения «смена собственника» семантические предикаты - *продавать* (лишаться собственности в пользу адресата за деньги), *покупать* (начинать иметь собственность от бенефицианта за деньги). Семантическое сообщение фактически является планом простого предложения и обычно описывается в терминах некоторого формализма, например, SPL [Kasper 1989]. Однако между семантическими сообщениями и предложениями текста нет однозначного соответствия, так как некоторые дискурсивные отношения между сообщениями реализуются внутри предложения, например, в виде причастного оборота, придаточного предложения или отглагольного имени с партиципантами.

В качестве примера семантического сообщения, записанного на языке SPL, приведем сообщение, порожаемое планировщиком системы AGILE:

```
(S / DIRECTED-ACTION
:LEX RISOVATJ
:ACTOR
(D / OBJECT :LEX POLJZOVATELJ)
:ACTEE
(L / OBJECT :LEX LINIJA )
)
```

Это сообщение поступает на вход грамматическому компоненту систему, на основе которого строится полностью реализованная грамматическая структура высказывания

«Пользователь рисует линию».

Виды структур планов текста, организующих сообщения

Независимо от того, является сообщение фрагментом предметной информации, которая только может быть оформлена лексическим предикатом, или это уже готовый план предложения с лексическим предикатом в вершине, сообщения организуются в структуру дискурсивными отношениями. В соответствии с природой сообщения и типом текста дискурсивные отношения имеют предметный или риторический характер, т.е. основываются на положении дел в реальности или вносятся автором текста в соответствии с коммуникативной задачей и типом генерируемого текста.

Простейшая структура плана текста соответствует **порядку следования событий**

реальности. Эта стратегия используется при генерации простейших отчетов из потока числовых данных, при генерации частей текстов, представляющих последовательность действий, например, в инструкциях.

План текста может быть организован тематически. Обобщение **тематической структуры** текста было предложено в [Hasan, 1978]. В этой работе введено понятие “generic structure”, под которым понимается жесткая тематическая схема, состоящая из “стадий”. Каждая стадия характеризуется единством темы и определенными лексико-грамматическими средствами, соответствующими контексту в рамках данной стадии. Этот подход использовался, в частности, в системе *Komet* [Bateman, Teich, 1995], которая порождала биографические справки о художниках на основании информации из БД. Например, следующий фрагмент текста представляет тематическую стадию “основная деятельность, основные произведения” (major activities, major works) в справочном тексте-отчете о биографии художника:

“Behrens’sU principal activities were architecture and industrial design. UHeU made electrical appliances and prototype flasks. UHeU built the high tension plant and the turbine factory for AEG in 1908-1910. UHeU built a housing for the workers of AEG in Henningsdort. UHeU created a number of monumental buildings, such as the administration building of Mannesmann in Duesseldorf and the German embassy in St. Petersburg.”

Стадии тематической схемы относительно независимы друг от друга и могут комбинироваться при изменении требований к выходному тексту. Например, тексты метеосводок, генерируемые системой FoG, могут иметь различную тематическую структуру в зависимости от того, кому они адресованы. В метеосводках для населения, наиболее значимым параметром является температура воздуха, она и располагается в начале тематической схемы текста, а в морских сводках для кораблей самой важной является информация о направлении и силе ветра, поэтому текст начинается с них.

План текста может организован в соответствии с **дискурсивной стратегией**, учитывающей структуру входных данных и коммуникативную цель текста. Дискурсивные стратегии отражают сложившуюся в данной культуре логику передачи определенного вида информации от автора текста к адресату. Термин “дискурсивная стратегия” был предложен К. Маккьюин и впервые реализован в ее системе TEXT в виде **риторических стратегий описания объектов** [Маккьюин, 1989]. Маккьюин

использовала в качестве входных данных БД по военным кораблям и рассматривала три коммуникативных цели: Определить, Описать или Сравнить.

Для описания дискурсивных стратегий Маккьюин использовала предикативные схемы, составленные в терминах риторических предикатов. *Риторические предикаты* характеризуют акты предикативности, которые может использовать говорящий. Они отражают структурное отношение между двумя пропозициями или фрагментами текста большего размера. Каждому риторическому отношению приписывалось одно из множества значений, таких, как *атрибутив, аналогия, состав, сравнение, тема, вывод, идентификация, переименование* и др.). *Предикативная схема* - это образец, определяющий, как должна быть построена последовательность изложения. В системе TEXT в виде схем были представлены четыре различных предикативных модели: *идентификации, состава, атрибутивности и противопоставления*. Как и правила определения содержания, предикативные схемы могут быть построены в результате тщательного анализа предварительно составленного корпуса текстов и обсуждений с экспертами в исследуемой ПО. Работа системы TEXT с использованием предикативных схем выглядела следующим образом: получая на входе коммуникативную цель (*определить, описать* или *сравнить*), система выбирала предикативную схему в соответствии с той информацией, которая имела в БД, и по выбранной схеме устанавливала, что должно быть сказано и в каком порядке. Ниже приведен упрощенный пример описания, построенного по схеме “состав” [Маккьюин, 1989] (в квадратных скобках в примере указаны риторические предикаты):

*“Паровые и электрические торпеды.
[состав:] Современные торпеды
бывают двух основных типов.
[глубинная идентификация (глубинный
атрибутив):] Модели с паровым
двигателем развивают скорость от 27
до 45 узлов и имеют радиус действия
от 4000 до 25 000 ярдов. [сравнение:
] Модели с электрическим двигателем
обладают близкими характеристиками,
[глубинная идентификация (глубинный
атрибутив):] но не оставляют
кильватерного следа, создаваемого
выбросом паровой турбины.”*

*Рис. 3 Пример описания,
построенного по схеме Состав.*

Кроме дискурсивных стратегий Маккьюин, отражающих структуру описаний объектов, в генераторах используются также **дискурсивные стратегии описаний пространства**, которые изучаются когнитивной лингвистикой. Они применяются для планирования содержания текстов, описывающих сцены или картины,

например, интерьер комнаты, девушка у окна и т.п. Но вместо риторических предикатов в этих планах организующими элементами являются не отношения, а физические и функциональные свойства предметов описываемого пространства в узлах структуры плана текста, например, массивность, центральность, вертикальность, автономность и т.п.

Расширением и углублением понятия риторических предикатов являются риторические отношения **Теории риторических структур** - Rhetorical Structure Theory (RST) [Mann and Thompson, 1987]. RST предлагает универсальный набор риторических отношений, которые организуют фрагменты ЕЯ текста с целью воздействия на адресата. Риторические отношения имеют определение, в которое входит зона "эффекта". Например, эффект отношения EVIDENCE между некоторыми главным и второстепенным фрагментами текста состоит в том, что доверие адресата к информации, представленной в главном фрагменте, увеличивается. Зона эффекта позволяет превратить определение риторического отношения в оператор, планирующий содержание текста в соответствии с заданной целью, что позволяет использовать подход к планированию содержания с помощью планирующих операторов, известный из Искусственного интеллекта.

План текста как **декомпозиция цели в терминах действий**. Планирующие операторы хорошо известны в приложениях ИИ [Sacerdoti 1977], в частности, в RST-планировщиках. Планирование здесь означает рациональную организацию действий для достижения поставленной цели. Цель раскрывается как последовательность действий, которые, в свою очередь, могут быть представлены как цели, и опять раскрыться как последовательность действий до достижения достаточно примитивных действий, которые могли бы быть непосредственно выполнены. Планирование предполагает три вещи: цель (проблема, которая должна быть решена), библиотека планов (набор операторов планирования, каждый из которых направлен на достижение определенной цели) и метод планирования (алгоритм). Каждый оператор планирования объединяет следующую информацию: эффект (состояние, принимающее истинное значение после применения плана, то есть цель), ноль, одно или более предусловий (условий, которые должны быть выполнены, чтобы план мог быть применен) и тело (набор действий, средств по достижению цели). Немного упрощенно, иерархический планировщик работает, производя рекурсивную декомпозицию коммуникативной цели текста с помощью планирующих операторов, предусловия которых выполнены и эффекты соответствуют целям, к которым применяются операторы. Описания систем, использующих такой

подход, могут быть найдены в [Novy 1993] и [Paris 1993].

Заключение

В статье перечислены разнообразные виды входных данных и планов текстов, которые используются в реальных системах ГЕЯ. Ситуация такова, что к настоящему моменту не выявлено четких принципов, как можно единообразно задавать входные знания и описывать стандартный процесс преобразований. Высказываются общие мнения о характере преобразований, в частности, в статье [Evans et al. 2002] говорится, что ГЕЯ характеризуется исключительно лингвистической направленностью всех проводимых ее рамках действий. Входные данные могут иметь и нелингвистический характер, но все действия по преобразованию этих данных должны быть лингвистически мотивированы и повышать уровень лингвистического представления информации. Таким образом, остается актуальной традиционная проблема генераторов, состоящая в том, что генераторы сильно ориентированы на конкретную задачу. С другой стороны, источники знаний создаются с помощью хорошо отлаженных формальных средств, но при этом содержащаяся в них информация не имеет продуманной структуры. С этим связаны проблемы, возникающие при использовании произвольных источников знаний в интеллектуальных приложениях, в частности и для целей генерации [Dale et al. 1998].

В докладе вопросы планирования будут иллюстрироваться экспериментами по генерации текстов в системе DEMLinG [Болдасов 2004].

Список литературы:

- 1) John Bateman and Michael Zock. Natural Language Generation, R. Mitkov (Ed.) Oxford University Press, Oxford 2001.
- 2) Жигалов В.А., Соколова Е.Г. 2001: InBASE: ТЕХНОЛОГИЯ ПОСТРОЕНИЯ ЕЯ-ИНТЕРФЕЙСОВ К БАЗАМ ДАННЫХ. Труды международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. т.2, стр. 123-135 Аксаково 2001.
- 3) Kruijff, G.-J., Teich, E., Bateman, J., Kruijff-Korbayova', I., Skoumalova', H., Sharoff, S., Sokolova, L., Hartley, T., Staykova, K. & Hana, J. (2000), 'A multilingual system for text generation in three slavic languages'. In Proceedings of the 18th Conference on Computational Linguistics (COLING 2000), Universita"t des Saarlandes, Saarbru"cken, Germany, pp. 474-480.
- 4) Иорданская Л. (1992) Коммуникативная структура и ее использование в системе текстовой генерации // Международный форум по информации и документации. Т. 17, №2. С. 15-28

- 5) Kasper, R. 1989. A flexible interface for linking applications to Penman's sentence generator. In Proceedings of the 1989 DARPA Speech and Natural Language Workshop, pages 153-158.
- 6) Hasan, R. (1978) Text in the Systemic-Functional model // in Dressler (ed.) "Current trends in text linguistics", de Gruyter, Berlin, pp. 228-246.
- 7) Bateman, J., Teich, E. (1995) Selective information presentation in an integrated publication system: an application of genre-driven text generation. Information processing & management, Vol. 31, No. 5, pages 753-767.
- 8) Макьюин К. Дискурсивные стратегии для синтеза текста на естественном языке // НЗЛ. Вып. XXIV. М., 1989. С. 311-356
- 9) Mann, W.C. and S.A. Thompson. 1987. "Rhetorical Structure Theory: description and construction of text structures", in: Kempen, G., (ed.) Natural Language Generation: Recent Advances in Artificial Intelligence, Psychology, and Linguistics. Kluwer Academic Publishers, Boston/Dordrecht, 85-96.
- 10) Sacerdoti E. 1977. A structure for plans and behavior. North Holland, Amsterdam.
- 11) Hovy, E. 1993. "Automated Discourse Generation using discourse structure relations". Artificial Intelligence, 63, 341-385.
- 12) Paris, C.L. (1993). User modelling in text generation. Frances Pinter, London.
- 13) Evans, R., P. Piwek and L. Cahill (2002). What is NLG? In: Proceedings of INLG02, New York, USA, 1-3 July 2002.
- 14) Robert Dale, Stephen J Green, Maria Milosavljevic, Cecile Paris, Cornelia Verspoor and Sandra Williams (1998) "The Realities of Generating Natural Language from Databases" In Proceedings of the 11th Australian Joint Conference on Artificial Intelligence. 13-17 July 1998. Brisbane, Australia.
- 15) Болдасов М.В. 2004: Разработка прикладных систем генерации типовых текстов на ЕЯ на основе представления информации на языке XML. Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции Диалог'2004, Москва: "Наука", стр. 55-61.