

УДК 004.7+004.8

К.С. Лащенко, Н.Г. Соколов, О.А. Гудаев
Донецкий национальный технический университет

ОНТОТЕКСТОЛОГИЯ ИНТЕРНЕТА

Аннотация.

К.С. Лащенко, Н.Г. Соколов, О.А. Гудаев.
Онтотекстология интернета. Представлена новая структура организации знаний и различного текстового материала в сети Интернет, которая позволит существенно улучшить использование памяти серверов путем разумного выбора текстового материала и удаления повторяющихся текстов среди множества сайтов.

Всем нам привычное «облако тегов» отражает общественное предпочтение по использованию слов в подборке текстов. Облако тегов строится на основе колоссальных затрат на память и вычисление селективной функции, которую выполняет компьютерная система. Как было бы просто отличать популярную книгу в библиотеке по лацкану обложки, натёртого до блеска, по сравнению с рядами книг, притрушенных пылью. Разветвлённая компьютерная сеть семантического Интернета собирает статистику хронологии обращения читателя к тексту. Препарируя эмоции, автоматическая функция конструирования «облако тегов» подменяет назначение ключевых слов текста. Допустим, автор текста явно выпишет ключевые слова. Автор может субъективно заблуждаться, что помог читателю понять текст. Генератор текста, расставляя акценты, далеко не поможет понять суть идеи. Понимание текста зависит от уровня развития мышления читателя или склонности его интересов. Медик с трудом поймёт текст по химии. Научно выражаясь: репрезентация получаемых из текста знаний отражается в «картине мира» читателя. Встраивание знаний осуществляется в модель мышления. Разные модели мышления химика и медика имеют

характер абсолютно разных взаимоотношений. Уместно предположить, что мышление – это глобальный процесс, а «картина мира» - это динамическая память мозга, изменяющийся, на лету, локальный процесс мышления. В итоге можно сказать, что «Облако тегов», автоматически порождённое нам в помощь, навязывает нам извне сюрреалистическую точку зрения, а ключевые слова автора текста недостаточно объективны. Автор текста не объективен: он знает о смысле текста всё, но не знает, как донести эту суть читателю, не имея представления о его модели мышления. На что же тогда опереться в понимании текста? Решение будет очевидно, если глубоко задуматься:

- о форме и структуре текста, как входных данных для анализа;

- о процессе написания текста автором;

- о скрытом, тайном смысле высказанной автором идеи, но не раскрытой до конца в тексте.

Самое трудное в поисках решения данной проблемы – это, донести до читателя скрытый, тайный смысл высказанной идеи, которую автор не смог раскрыть в тексте до конца. Таким образом, получается ситуация «детерминированного хаоса в квадрате»: хаос модели знаний читателя о рассматриваемом вопросе в тексте, перемноженный на хаос изложения автором сути идеи, в виде формального текста. В вопросе «скрытых» смыслов существует две гипотезы.

Первая гипотеза заключается в следующем: первоисточник зарождения мысли пытается построить объёмную картину мира рассматриваемого предмета, оперируя простыми понятиями и устанавливая максимальное количество реферативных связей на момент написания текста. Достоинство подхода, изложенного в гипотезе – это понятность. Недостаток – это устаревание реферативных ссылок, в чём и заключается исторический процесс эволюции идеи, заложенной в тексте.

Вторая гипотеза: эволюция первоисточника протекает не в исторической хронологии, а в логической.

Уточним, данная гипотеза предполагает, что первоисточник может эволюционировать в терминах классической логики, а

может «нарушать симметрию» и эволюционировать по правилам неклассической логики. Поясним «нарушение симметрии»: не диалектический материализм; не развитие по спирали; не выполнение закона отрицания отрицаний.

Такая сложная ситуация с неклассической логикой развития первоисточника идеи может называться «информационной волной». Здесь стоит сразу разобраться и получить ответ на вопрос «Почему процесс не “исторический”?» Ожидается, что высказанная идея неизменна во все века, а исторически только уточняется и раскрывается новыми прецедентами применения на практике. То есть имеет хронологию применения, а суть идеи остается неизменной, пройдя сквозь века. Но это не так!

Тогда возникает еще несколько вопросов. Как же происходит трансформация идеи на самом деле? И кто привносит вклад в эти нарушения? Процесс эволюции «живой»! Чередование формализации писателя и интерпретации читателя приводит к разрыву первоисточника с сутью происходящего в современности, в виду желания устанавливать новые реферативные ссылки. Ссылки имеют субъективную окраску, так как их устанавливают новые исследователи текста.

Требование идти со временем «в ногу» заставляет экспериментаторов пересказывать идею первоисточника в понятных современнику терминах, ссылаясь на существующие прецеденты и теории. Тогда искажения неизбежны. Так возникает «информационная волна». Например: «писк» - это противный скрип лебёдки или модный тренд, вызывающий воодушевление.

Можно предположить, что сама современность привносит искажения в первоисточник. При смене хронологии событий искажение изначального изложения онтологий сильнее раскрывают их суть непосредственно для современных читателей. Текст будет обрастать все новыми и новыми искажениями. Прохождение онтологий через хронологию событий можно наблюдать на рисунке 1. В какой-то определенной хронологической последовательности, современный автор может интерпретировать текст добавляя

новые или отнимая старые факты. Таким образом, имеем искажения полученные на основе прецедентов.



Рисунок 1 – Трансформация онтологий текста

Почему происходит искажение смысла первоисточника мысли в ходе передачи информационной волны? А может это новые знания, не сопоставимые с источником. Так происходит из-за сущности формы изложения текста на естественном языке. Форма изложения текста подчинена главному постулату: аргументы и факты облакаются в эмоциональный контекст, чтобы быть интересным читателю. Нудный и монотонный слог усыпляет сознание и блокирует размышления над прочитанным. Человек на подсознательном уровне начинает блокировать получаемую информацию, считая ее угрозой или бесполезным засорением памяти. Даже если сама суть идеи важна и интересна читателю, он не сохранит и не поймет ее из-за неправильной подачи автором. В ходе передачи информационной волны удачный художественный контекст превалирует над логикой излагаемых фактов. Пересказчик интерпретирует понравившийся текст, по своему мнению, вкладывая в контекст новое звучание, превалирующее над первоначальным замыслом, но это не гарантирует мотивированного понимания идеи всеми читателями. Все люди разные, что интересно одному, другой сочтет за мусор.

Поэтому, перед пересказчиком текста стоит практически нереализуемая задача – донести суть идеи до всех читателей. Так бывает, что перевод с японского в несколько раз искуснее оригинала.

Как же найти предел психологического барьера «скучно – не понимаю» / «нравится – всё понятно!»?

Текст, проходя через субъективные точки зрения, нелинейно трансформируется. Это реалии жизни помогают всему

человечеству «не закиснуть». Может ошибаться пересказчик? Да. Его семантическая ошибка в тексте принадлежит иррациональному миру. Но иррациональный мир – источник новых знаний [1, с. 225]. Как делать полезные ошибки? Ответ: фантазия и воображение. Молодая теория воображения не может дать рекомендации на все приёмы художественного оформления текста [2]. Но кое-что может сказать о форме и структуре текста. В компьютерном мире Интернета существует простая формула: контент – это текст плюс контекст. В качестве скрытых тегов выступают ключевые слова электронных документов, заданные для SEO-оптимизации, и автоматически подобранные поисковой машиной аннотированные ссылки. Теория воображения рассматривает процесс создания текста. Поэтому, косвенно затрагивает литературную структуру упаковки смысла в текст. Например, моделью структуры будет <Пейзаж, Персонаж, Правила жизни, Сцена>. Сцена – это полигон, где персонажи проживают явления или ситуации по своим правилам жизни в заданной локации. Структура без формы «рассыплется». В логике воображения не рассматривается низкоуровневая форма формализации художественного замысла. Может быть – это путь развития молодой науки.

Рассмотрим формы изложения онтологий в виде текста. Эти формы не премудрости копирайта. Онтологии «живут» в некоторой форме изложения текста и программируют смысл прочтения. Для конкретного психологического типа читателя устанавливается баланс между любимой формой изложения, её должно быть побольше, и понятной, которая не должна раздражать, а приятно удивлять, её должно быть не много. Стоит перечислить все формы изложения текста по категориям в таблице 1.

Таблица 1 – Формы изложения текста

Форма	Категория	Конкретизация
Имя, название, безусловный адрес, именование	Формальная	Таксономия, термин
Перечень целей,	Формальная	Мотивация:

целевая мотивация, назначение		{принципы, законы}, инструкция, приказ
Перечисление результатов	Взаимосвязь с другими объектами	Модель черного ящика, через входные и выходные данные, плоды, влияние на внешнюю среду
Область применения	Взаимосвязь с другими объектами	Использование, условия внедрения
Функционирование	Процедурная	Порождающая процедура, схема функционирования, суперпозиция функций
Рецепт, сценарий поведения, алгоритм синтеза	Процедурная	Организация процесса, логистика, практический опыт
Идеальная модель	Структурная	Синтез, идеализация через обобщение
Аналогия	Структурная	Прообраз, искусственное сочетание
Перечисление свойств, характеристик	Структурная	Косвенная адресация, ассоциативный ключ
Множество вопросов	Художественная	Интервьюирование
Отрицание свойств	Художественная	Отсечение известных знаний
Метафорический образ	Художественная	Мифотворчество, мистификация, иносказательный образ

Форма – это мощный инструмент программирования извлечения смысла из текста читателем. Форма принимает вид опосредованной интерактивной игры типа FEC, без возможности запроса на повторную передачу смысла текста в другой форме.

Если писатель не играет с читателем, то его текст скучен и монотонен, что в свою очередь включает защитный механизм читателя воспринимать информацию как враждебную. Писатель должен уметь создавать разнообразные формы текста. Компьютерные чат боты должны уметь различать все формы текста, чтобы пройти тест Тьюринга. Например, название текста в форме абсолютной адресации «Lexicode for Augmented Reality» вызовет бурю эмоций у передовых учёных и текст надо будет писать впервые [3].

Чем более схожи идеи, тем они дороже как мнение. Коммерческая модель онтологического Интернета – это когда плата берётся за внесение в хранилище. Анализ текста и поиск схожести будет осуществлять данная модель Интернета, в чем проявляется ее несомненное преимущество над современной моделью. Пример адресной строки приведен на рисунке 2.

Поисковый запрос и представление запроса на естественном языке приведено ниже:
 Я. {Олег}. Наука. {Коды. [телекоммуникацион.]} | Публикации. {статьи}. Поиск. {маркеры}. «Научные статьи Олега по теме телекоммуникационные коды. Ключ для поиска - “маркеры”».

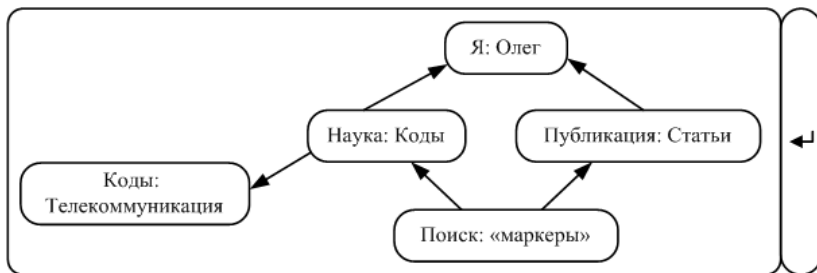


Рисунок 2 – Адресная строка

Наслоение компьютерных фактов теряется в большом количестве сведений. Любая система накопления зависит от желания сохранить всё на потом. Репозиторий данных разрастается в необъятное хранилище. Хранение контекста добавляет широту обзора сведений, но это ведёт к ещё большему

разрастанию данных. Избавится от контекста невозможно. Вне контекста сведения малозначительны. Легкость, с которой формируются слои доказательных материалов, делает затруднительным поиск истины. Контекст для несвязанных фактов страдает многократным повторением одного и того же. Автор не умеет выделить главное и внести это правильно в виде контекста в текст. Мыслить многомерным миром онтологий только для того, чтобы дополнить сообщение контекстом, автору идей кажется невероятной работой с субъективным конечным результатом. Контекст нужен только в будущих периодах использования! Поэтому, автор современного Интернета, не мотивирован и считает работу по созданию текста с онтологическим контекстом бессмысленной. Для примера важности контекста, составим граф развития гипотезы (см. рис. 3). Имя A1 – это автор, первоисточник мысли, который выдвинул гипотезу M1. Авторы: {A1, A2, A3, A4}. Идеи: {M1, M2, M3, M4, M5, M6, M7}.

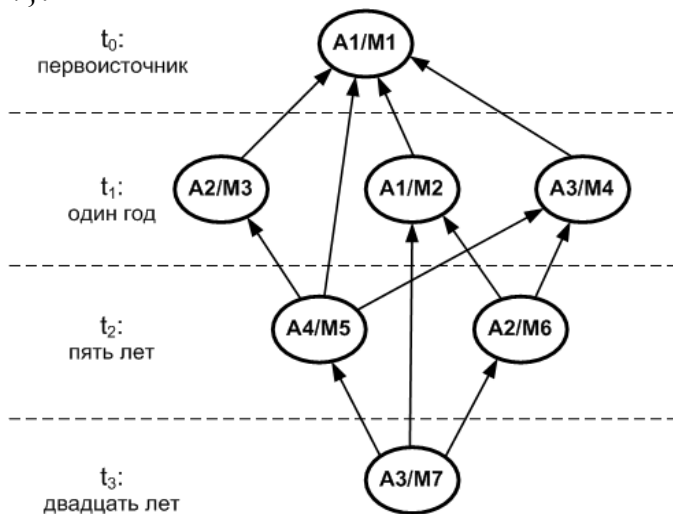


Рисунок 3 – Граф развития гипотезы

Чтобы не засорять хранилище однотипными текстами, плата за текст схожий с уже имеющимися текстами берется намного

выше, чем если бы текст представлял собой большую уникальность по сравнению с остальными в хранилище. Это будет сильнее мотивировать автора к написанию более уникальных текстов и экономить огромное количество пространства.

Выводы. Таким образом, данное представление организации Интернета позволяет создать более связную по смыслу онтологическую систему, которая бы могла учитывать различные формы изложения текста для одной и той же идеи. Такая подача информации позволит донести до любого читателя суть идеи ввиду наличия огромного множества форм изложения текста. В свою очередь у авторов или пересказчиков текста сформируется своя аудитория читателей, а у читателей сформируется ряд любимых авторов, понравившихся по форме изложения идей в тексте. Ввиду вышесказанного, существенно повысится мотивация по созданию доступного изложения любой идеи, мысли, гипотезы. В результате компьютерная система онтологической организации Интернета сможет: переформулировать текст, организовывать обратную связь вовлеченных авторов, вопреки современной ленточной подачи параллельных комментариев, осуществлять создание исторической хронологии всегда связанных первоисточников.

1. Ракилов, А.И. Философия компьютерной революции [Текст] / А.И. Ракилов. – М. : Политиздат, 1991. – 287 с. – ISBN 5–250–01308–2.

2. Шрагина, Л.И. Логика воображения [Текст] : Учебное пособие / Л.И. Шрагина. – 2-е изд., дороб. – М. : Народное образование, 2001. – 192 с.

3. Гудаев, О. Иероглификация вселенского масштаба [Текст] / О. Гудаев // Наука и техника. – 2014. – № 7(98). – С. 7–10.