

ЗАЛИЧЕВ Николай Николаевич - доктор технических наук, профессор, главный редактор межотраслевого научно-технического журнала «Интеграл»

МАЧКИН Петр Иванович - советник Международного Конгресса промышленников и предпринимателей

СЕМАН-ТЕХНОЛОГИЯ И ЕЕ ПРИМЕНЕНИЕ В СИСТЕМЕ ЭЛЕКТРОННЫХ БИБЛИОТЕК

Сущность и содержание СемАн-технологии

СемАн-технология (дословно: технология Семантического Анализа) представляет собой по своей сущности, содержанию и реализуемым функциям программно-информационную реализацию методологии семантического (смыслового) анализа и энтропийной оценки обрабатываемой, с использованием средств электронной вычислительной техники, текстовой информации, описывающей интересующую пользователя какую-либо конкретную сферу человеческой деятельности и соответствующую ей область знаний или группу областей знаний.

СемАн-технология предназначена для проведения с использованием средств электронной вычислительной техники семантического анализа и энтропийной оценки обрабатываемой исходной текстовой информации, описывающей анализируемую пользователем какую-либо конкретную сферу человеческой деятельности и соответствующую ей область знаний или группу областей знаний.

Принципы, способы и методы построения и обеспечения функционирования СемАн-технологии позволяют применять ее в составе автоматизированных систем различного типа и назначения для решения широкого круга практических задач и в том числе в составе автоматизированных систем электронных библиотек.

Целью СемАн-технологии является построение с использованием средств электронной вычислительной техники объективного прогноза развития существующих знаний и формирования новых знаний в какой-либо анализируемой пользователем данной технологии конкретной сфере человеческой деятельности и в соответствующей ей области знаний или в группе областей знаний.

Принципы, способы и методы реализации семантического анализа и энтропийной оценки обрабатываемой исходной текстовой информации позволяют построить на их базе основные элементы СемАн-технологии - программно-информационные комплексы различной информационно-вычислительной мощности и назначения по проведению семантического анализа и энтропийной оценке обрабатываемой текстовой информации (книжной, журнальной, газетной и т.п., описывающей анализируемую пользователем данной технологии какую-либо конкретную сферу человеческой деятельности), ориентированные (skonфигурированные или «заточенные») на обработку информации в каких-либо конкретных сферах и областях человеческой деятельности (и в соответствующих им областях знаний, или в группах областей знаний) и предназначенные для построения объективного прогноза развития существующих знаний и формирования новых знаний в этих исследуемых пользователем сферах человеческой деятельности и в соответствующих им областях знаний или в группах областей знаний.

Базовой основой построения программно-информационных комплексов СемАн-технологии различной информационно-вычислительной мощности и назначения является созданная авторами данной технологии оригинальная методология, математический аппарат, а также разработанные на их основе и апробированные на практике алгоритмы проведения семантического анализа и энтропийной оценки обрабатываемой тестовой информации.

В СемАн-технологии реализованы принципиально новые приемы, способы, методы и алгоритмы семантической обработки и энтропийной оценки текстовой информации, построенные не на принципах искусственного интеллекта или нейросетей, - в ее основу заложены результаты проведенных системных исследований процесса мышления человека, а также природы информации и закономерностей ее проявления на семантическом (смысловом), а не на более упрощенном синтаксическом (количественном) уровне.

Особенности и преимущества СемАн-технологии

Программно-информационные комплексы СемАн-технологии существенно отличаются и имеют свои особенности и преимущества перед широко применяемыми в настоящее время в составе автоматизированных систем (АС) различного типа и назначения синтаксическими информационно-аналитическими комплексами и системами. Эти особенности и преимущества обусловлены различием между синтаксическим и семантическим методами анализа обрабатываемой информации.

И без уяснения и осмысления основного, базового отличия синтаксического и семантического методов анализа обрабатываемой информации невозможно понять сущность и содержание механизма реализации в СемАн-технологии (построенной на основе семантического метода анализа обрабатываемой информации) прогноза

развития какой-либо анализируемой области знаний (или сферы человеческой деятельности), а также синтеза новых знаний в этой анализируемой сфере человеческой деятельности.

Главное отличие синтаксического и семантического методов анализа обрабатываемой информации заключается в различной по своей сущности мере количества информации, используемой в этих методах.

В синтаксическом методе анализа обрабатываемой информации ее мерой, введенной К. Шенноном, является количество знаков (или слов, составленных из этих знаков, предложений, составленных из входящих в эти предложения слов, и прочих параметров, количественно определяющих и описывающих каким-либо способом обрабатываемую информацию), и измеряемой количественно в битах (двоичных единицах) этой информации.

В семантическом методе анализа обрабатываемой информации (практически реализованном в СемАн-технологии) ее мерой является количество элементарных семантических единиц, отражающих обрабатываемую информацию, и измеряемой количественно в уровне энтропии этой информации.

Синтаксическая мера информации совершенно не подразумевает смысловую обработку информации, так как ей это не нужно по ее определению. Семантическая же мера информации не только подразумевает, но и обязывает работать со смыслом, заложенным в «синтаксисе» обрабатываемой информации: в словах, предложениях, их взаимном пересечении и т.д., и ей это необходимо по ее определению.

И именно для того, чтобы выявить в явном и формализуемом виде смысловое содержание обрабатываемой информации, при создании СемАн-технологии потребовалось ввести три новых базовых понятия, позволивших разработать необходимый математический аппарат и построить на его основе полностью и строго формализуемый процесс семантического (смыслового) анализа обрабатываемой информации [1]:

- 1) **элементарная семантическая единица (ЭСЕ)**, представляющая собой законченную мысль в виде утверждения, являющаяся фактически семантической мерой количества обрабатываемой информации, и выражающаяся количественно в уровне энтропии этой информации;
- 2) **информационный фантом (ИФ)**, представляющий собой информационное отражение фрагмента анализируемой реальности, описываемый обрабатываемой информацией и состоящий из «мозаики» элементарных семантических единиц;
- 3) **квант снижения стохастичности (КСС)**, представляющий собой минимальную величину изменения интегральной плотности распределения истинности элементарных семантических единиц в составе сформированного информационного фантома фрагмента анализируемой реальности в результате проведения семантического анализа и энтропийной оценки обрабатываемой информации.

Семантический анализ обрабатываемой информации осуществляется в СемАн-технологии путем сопоставления детерминированной (т.е. симметричной реальности) и случайной (асимметричной) составляющих информационного фантома с использованием при этом КСС как инструмента оценки степени «сближения» этих двух составляющих ИФ.

Именно эти три новых базовых понятия семантического метода анализа обрабатываемой информации и разработанный на их основе математический аппарат позволяют реализовать на практике, во-первых, полностью и строго формализованный процесс семантического (смыслового) анализа обрабатываемых данных, во-вторых, осуществить прогноз развития анализируемой области знаний (сферы человеческой деятельности) и, в-третьих, сформировать новые знания, то есть осуществить синтез новых знаний в этой анализируемой сфере человеческой деятельности.

Предложения и рекомендации по расширению функциональных характеристик автоматизированных систем электронных библиотек с использованием СемАн-технологии

Электронная библиотека любого уровня организации и обеспечения функционирования (корпоративная, ведомственная, межведомственная, общегосударственная, межгосударственная и т.п.) в самом общем виде может быть представлена как совокупность двух основных ее элементов: электронных фондов (структурированных массивов) разнородных данных (текстовых, символьных, графических, а также аудио- и видеоинформации) и каналов и сетей связи и передачи данных между ними. При этом совокупность этих элементов объединяется и структурируется в рамках каких-либо территориально-распределенных автоматизированных систем, а для обеспечения их функционирования создается необходимая человеко-машинная среда взаимодействия пользователей электронных библиотек с их ресурсами.

Расширение функциональных характеристик автоматизированных систем электронных библиотек возможно на основе широкого использования при построении и обеспечении их функционирования программно-информационных комплексов СемАн-технологии различной информационно-вычислительной мощности и назначения путем практической реализации для этого следующих мероприятий:

- 1) включения в состав данных АС программно-информационных комплексов СемАн-технологии в качестве функциональных комплексов этих систем;
- 2) настройки программно-информационных комплексов СемАн-технологии на предметно-функциональную ориентацию их функционирования в составе данных АС в рамках автоматизации конкретных сфер

практической деятельности конкретных должностных лиц, в соответствии с их направлениями деятельности и должностными обязанностями.

Программно-информационные комплексы СемАн-технологии различной информационно-вычислительной мощности и назначения при их применении в составе автоматизированных систем различного типа и назначения и в том числе в составе АС электронных библиотек будут способны пропускать через себя огромные массивы обрабатываемой в этих системах текстовой информации, а заложенная в эти комплексы технология проведения семантического анализа (ориентируясь на смысловое содержание анализируемой информации) и энтропийной оценки этой исходной синтаксической информации позволят решать следующие задачи конечного пользователя:

- 1) выявлять информацию, не содержащуюся фактически в явном виде в каждом отдельно взятом источнике исходной информации, но появляющуюся в результате комплексного семантического анализа и энтропийной оценки совокупности различных (разнородных) источников анализируемой информации;
- 2) осуществлять объективный прогноз развития существующих знаний в какой-либо анализируемой области знаний или группе областей знаний;
- 3) формировать новые знания в какой-либо анализируемой области знаний или группе областей знаний;
- 4) предсказывать скачки в развитии какой-либо анализируемой области знаний или группе областей знаний, то есть переход количественных изменений в новое качественное состояние в этой области знаний или группе областей знаний;
- 5) выявлять ложные пути в развитии анализируемой области знаний или группе областей знаний и разрабатывать предложения (путем формирования новых знаний) по предотвращению дальнейшего ее развития по выявленным ложным путям.

СемАн-технология в настоящее время теоретически полностью разработана, и на некоторых примерах ее применения получено подтверждение и объективное доказательство практической реализуемости заложенных в основу ее построения методологии, математического аппарата и алгоритмов семантического анализа и энтропийной оценки обрабатываемой информации.

Но полнофункциональные программно-информационные комплексы СемАн-технологии различной информационно-вычислительной мощности и назначения для их практического применения в различных видах и сферах человеческой деятельности на настоящий момент времени пока еще не созданы. Это задача ближайшего будущего. Залогом же того, что СемАн-технология может быть создана, являются, как указано выше, результаты практического подтверждения ее реализуемости.