

Гипертекст и ультратекст

А.В. Чечкин

По гипертексту формируется оперативная экспертная система на базе понятий и методов математической информатики. Оперативная экспертная система воспринимает данный гипертекст и генерирует новый путем дополнения исходного гипертекста его логическими следствиями на основе имеющихся в экспертной системе знаний. Новый сгенерированный гипертекст называется ультратекстом.

1. Идея ультратекста

Понятие гипертекста широко используется в научной литературе, технической документации и т.п. [3].

Гипертекстом называется ориентированный граф (сеть), вершины которого нагружены словами в некотором алфавите, в частности текстами на естественном языке. Тем самым гипертекст есть частично упорядоченный набор своих частей в отличие от обычного текста, все части которого линейно упорядочены.

Примерами гипертекста являются: научные, технические и т.п. документации, библиографические и др. каталоги, справочные материалы и многое другое. По существу, к гипертексту относится семантическая структура любого распространенного предложения естественного языка, любого повествовательного текста на естественном языке.

Идея ультратекста состоит в следующем.

- Дан гипертекст, т.е. ориентированный граф, вершины которого нагружены данными (сведениями) об объектах. Направленная дуга гипертекста — это ссылка (связка) от одних данных к другим, отражающая связи между объектами.
- По гипертексту формируется оперативная экспертная система, ориентированная на восприятие данного гипертекста.
- Оперативная экспертная система загружается данным гипертекстом.
- Оперативная экспертная система генерирует новый гипертекст путем дополнения исходного гипертекста его логическими следствиями на основе имеющихся в экспертной системе знаний.

Таким образом новый гипертекст есть результат осмысленного прочтения данного гипертекста соответствующей ему оперативной экспертной системой. Новый сгенерированный гипертекст называется *ультратекстом*.

2. ОПЕРАТИВНАЯ ЭКСПЕРТНАЯ СИСТЕМА (ОЭС)

ОЭС будем определять на базе понятий математической информатики [1]. В рамках идеи ультрамедиа ОЭС — это U -радикал [2].

ОЭС представляет собой сеть накопителей и преобразователей информации. Тем самым ОЭС является распределенной базой данных и знаний. При этом накопители информации понимаются как локальные базы данных (ЛБД), а преобразователи информации понимаются как локальные базы знаний (ЛБЗ).

Информация в ОЭС содержится главным образом в виде данных, распределенных по накопителям. Знания в ОЭС содержатся главным образом в преобразователях в виде правил, продукций, процедур и т.п.

Таким образом ОЭС является сетью информационных устройств двух типов: одни ориентированы на хранение, другие — на преобразование локальных данных.

Математической моделью ОЭС является ультрасеть [1]. Ультрасеть есть расширение сети Петри. В ней предусматриваются два вида разметки: разметка запросов и разметка ответов. Эти разметки изменяются по специальным правилам [1]. Тем самым ультрасеть моделирует информационные процессы в соответствующей ей экспертной системе.

Кроме разметок ультрасеть имеет математические нагрузки накопителей и преобразователей в виде соответствующих математических моделей. Для накопителей это — *ультрамножества*, а для преобразователей — *ультраоператоры* [1, 2].

Дадим необходимые определения.

Определение 1. Ультрасетью называется двудольный ориентированный мультиграф $G = (Q, S, U)$, где $Q = \{q_1, \dots, q_m\}$ — множество вершин, называемых позициями, $S = \{s_1, \dots, s_n\}$ — множество вершин, называемых переходами, U — множество дуг, соединяющих вершины только разноименных множеств Q и S . При этом множества Q и S не пересекаются. Множество U можно задать двумя матрицами инцидентности как в сетях Петри. Будем изображать графически позиции кругами, а переходы прямоугольниками.

Определение 2. Разметкой запросов (разметкой первого типа) в ультрасети $G = (Q, S, U)$ называется функция $f(x)$ отображающая множество всех вершин (и позиций, и переходов) во множество $\{0, 1\}$. Число $f(q_i)$ изображается на графе наличием или отсутствием знака плюс "+" в соответствующем q_i круге. Аналогично число $f(s_j)$ изображается на графе наличием или отсутствием знака плюс "+" в соответствующем s_j прямоугольнике. Процесс изменения разметки запросов в ультрасети называется функционированием первого типа. Это функционирование происходит от начальной разметки по следующим правилам:

- если переход s_j имеет нулевую разметку запросов, т.е. там нет знака плюс, но хотя бы в одной из его выходных позиций имеется ненулевая разметка запросов, то происходит обязательное срабатывание этого пе-

рехода. В результате срабатывания переход приобретает ненулевую разметку запросов и все его входные позиции приобретают или сохраняют ненулевую разметку запросов. При этом разметка запросов во всех его выходных позициях остается неизменной;

- если два или более перехода могут одновременно сработать, то их срабатывание независимое и осуществляется в любой последовательности или параллельно;
- срабатывание переходов продолжается до тех пор, пока может сработать хотя бы один переход. После чего ультрасеть останавливается. Итак, в результате функционирования первого типа в ультрасети знаками "+" выделяется подсеть (часть G), а именно (Q^+, S^+, U^+) .

Определение 3. Разметкой ответов (разметкой второго типа) в ультрасети $G = (Q, S, U)$ называется функция $g(q_i)$, отображающая множество позиций Q^+ , имеющих знаки плюс, т.е. ненулевую разметку запросов, во множество натуральных чисел с нулем. Число $g(q_i)$ изображается на графе ультрасети маркерами в круге q_i как в сетях Петри, при этом знак плюс в круге q_i аннулируется (маркер сильнее плюса).

Функционированием второго типа для ультрасети называется процесс изменения разметки ответов по той подсети ультрасети, которая имеет ненулевую разметку запросов по правилам сетей Петри. При этом

- после срабатывания очередного перехода, помеченного плюсом, происходит аннулирование знака плюс в этом переходе;
- после изымания маркеров из входных позиций там знак плюс не восстанавливается;
- маркеры появляются только в тех выходных позициях срабатывающего перехода, которые помечены знаком плюс, при этом появление маркера в позиции сопровождается аннулированием знака плюс в этой позиции.

Определение 4. В ультрасети $G = (Q, S, U)$ функционирование первого и второго типов происходит одновременно. При этом соблюдается правило "разметка ответов сильнее разметки запросов", что означает следующее:

- появление маркера в позиции (при функционировании второго типа) аннулирует знак плюс в этой позиции (маркер сильнее плюса);
- после срабатывания перехода (при функционировании второго типа) в переходе знак плюс аннулируется;
- после каждого срабатывания какого-либо перехода (при функционировании второго типа) в ультрасети аннулируются знаки плюс во всех входных позициях перехода и в самом переходе, если у этого перехода нет ни одной выходной позиции со знаком плюс.

Итак, в ультрасети сначала происходит выделение некоторой ее части (подсети) с помощью разметки запросов (плюсов), а затем только в этой части ультрасети происходит ее сокращение при помощи разметки ответов.

Правила сетевого моделирования ОЭС. Как мы уже отмечали, в ОЭС

имеются накопители (ЛБД) и преобразователи (ЛБЗ) информации. Сформулируем правила моделирования ОЭС ультрасетями.

Правило 1. Каждому накопителю поставим во взаимно однозначное соответствие неразмеченную позицию ультрасети.

Правило 2. Каждому преобразователю поставим во взаимно однозначное соответствие переход ультрасети. Очевидно что связь преобразователя с его входными и выходными накопителями моделируется соответствующими инцидентными этому переходу дугами.

Правило 3. Запрос будем моделировать ненулевой разметкой запросов (разметкой первого типа) в соответствующей позиции. При этом знак плюс "+" в позиции будет обязательно окрашен, т.е. снабжен записью (что спрашивается, о чем спрашивается и т.д.).

Правило 4. Ответ на запрос будем моделировать маркером ".", т.е. ненулевой разметкой ответов (разметкой второго типа) в соответствующей позиции с ненулевой разметкой запросов. При этом маркер "." в позиции будет обязательно окрашен, т.е. снабжен записью конкретного ответа на запрос.

Правило 5. Распространение запросов происходит следующим образом:

- начальный запрос приходит в некоторый накопитель или в несколько накопителей от пользователя ОЭС и в соответствующей позиции или нескольких позициях проставляется знак "+" с соответствующей окраской (записью);
- если позиция ультрасети имеет знак "+", т.е. если в накопитель пришел запрос, то в этом накопителе идет поиск ответа. При наличии в накопителе (ЛБД) нужных для ответа сведений, выдается ответ на запрос. При этом в позиции знак "+" меняется на маркер ".". Этот маркер должен быть окрашен конкретной записью ответа;
- если в накопителе, куда пришел запрос, нет всех сведений или части сведений для формирования ответа, то знак "+" вместе с его окраской в позиции сохраняется. При этом запрос продвигается к тому преобразователю (ЛБЗ), который имеет данный накопитель одним из своих выходных накопителей. Другими словами запрос продвигается по дуге против стрелки к соответствующему переходу ультрасети. В соответствующем этому переходу прямоугольнике проставляется знак "+", говорящий о том, что данный преобразователь вызван для решения соответствующей информационной задачи;
- если накопитель, откуда продвигается запрос, связан с несколькими переходами, для которых этот накопитель является выходным, то продвижение запроса происходит по определенной стратегии. Эта стратегия формулируется вне рамок формализма ультрасетей. Возможные стратегии продвижения запроса, например, следующие: одновременное продвижение запроса ко всем возможным преобразователям или попеременное продвижение запроса сначала к одному преобразователю, и только в случае

невозможности получения ответа через него, продвижение запроса к другому преобразователю и т.д.;

- если преобразователь имеет ненулевую разметку запросов, т.е. если он вызван для решения соответствующей информационной задачи, то он решает эту задачу. Если при этом преобразователь получит ответ на запрос, то ответ посылается в соответствующий выходной накопитель. При этом знак "+" в преобразователе (переходе) аннулируется, а в накопителе (позиции) знак "+" меняется на маркер ".", соответственно окрашенный записью ответа;
- если преобразователь, вызванный для решения информационной задачи, не может получить ответа на запрос без некоторых данных из его входных накопителей, то этот преобразователь возбуждает запросы к своим входным накопителям. Тем самым запрос продвигается по дугам от преобразователя к его входным позициям против стрелок и в этих позициях проставляются знаки "+", соответственно окрашенные конкретным текстом запроса;
- процесс распространения запросов выделяет знаками "+" подсеть ультрасети из накопителей и преобразователей ОЭС. При этом в некоторых накопителях появляются ответы, т.е. знаки "+" заменены знаками ".". Этот процесс происходит до тех пор, пока может происходить;
- если появится преобразователь, отмеченный ненулевым знаком запросов, у которого входные накопители сменили знаки "+" на знаки ".", и нет ни одного входного накопителя со знаком "+", то начнется обратный процесс получения ответов. При этом процесс распространения запросов прерывается;
- если процесс распространения запросов в ультрасети закончился в силу ее ограниченности и нет необходимых маркеров для обратного процесса, то функционирование ультрасети заканчивается. В этом случае по терминальным разметкам запросов видно, какие сведения требуется закачать в ультрасеть для организации в ней обратного процесса распространения ответов.

Правило 6. Распространение ответов происходит следующим образом. По выделенной знаками "+" части ультрасети будет происходить смена разметки запросов на разметку ответов по правилу 4. При этом с каждым разом выделенная знаками "+" подсеть будет сокращаться, так как знаки "+" аннулируются. Этот процесс получения ответов происходит до тех пор, пока может сработать хотя бы один преобразователь, отмеченный знаком "+".

Правило 7. Если в результате распространения ответов маркеры "." ответов пришли во все позиции, где были начальные запросы, то функционирование ультрасети заканчивается. Если нет, то возобновляется распространение запросов по правилу 5 и т.д.

Математическая нагрузка ультрасети. Обсудим сначала математическую нагрузку накопителя (позиции ультрасети), а затем преобразователя

(перехода ультрасети) информации ОЭС.

Определение 5. Математической нагрузкой неразмеченной позиции (накопителя) ультрасети называется множество математических моделей всех возможных сведений для этого накопителя. Такое множество носит название *ультрамножества* [1, 2].

Ультрамножество над множеством X — это декартово произведение опорного множества X моделей объектов предметной области и двух решеток L и P , которыми оснащается опорное множество. Решетка L называется решеткой понятий. L является булевой алгеброй с максимальным элементом $p = X$, являющимся самым общим понятием в L и с минимальным элементом $p = 0$, являющимся пустым понятием в L . Решетка P называется решеткой достоверностей. Она характеризует источники сведений об объектах опорного множества. P является гейтинговой алгеброй с максимальным элементом $l = 1$, понимаемым как истина и с минимальным элементом $l = 0$, понимаемым как ложь. Будем обозначать ультрамножество над X через [ULTR- X]. Итак

$$[\text{ULTR-}X] = P \times L \times X$$

Определение 6. Математической нагрузкой позиции ультрасети, отмеченной маркером "." (ответом) называется математическая модель сведения из соответствующего ультрамножества, в которой все части являются известными (постоянными). Математической моделью сведения является конкретный элемент $(p)l(x)$ из [ULTR- X].

Определение 7. Математической нагрузкой позиции ультрасети, отмеченной знаком "+" (запроса), называется математическая модель сведения из соответствующего ультрамножества, в которой некоторые части модели являются неизвестными (переменными). Тем самым математической нагрузкой знака запроса является элемент из [ULTR- X] вида $(p)?(x)$, или $(?)l(x)$, или $(p)l(?)$, или $(?)?(x)$ и т.д., который является окраской этого знака запроса.

Определение 8. Математической нагрузкой неразмеченного перехода (преобразователя) ультрасети называется математическая модель преобразователя информации, соответствующего этому переходу согласно правилу 2. Этой моделью является так называемый *ультраоператор* [1, 2].

Ультраоператором над оператором $C : X \rightarrow Y$ называется отображение ультрамножества [ULTR- X] в ультрамножество [ULTR- Y]. Будем обозначать ультраоператор через [ULTR- C], итак

$$[\text{ULTR-}C] : [\text{ULTR-}X] \rightarrow [\text{ULTR-}Y].$$

При этом, если $[\text{ULTR-}X] = P \times L \times X$, $[\text{ULTR-}Y] = Q \times M \times Y$, то

$$[\text{ULTR-}C] : (p)l(x) \rightarrow (q)m(y),$$

где $y = Cx$. В [1] доказывается, что ультраоператор задается тройкой отображений

$$[\text{ULTR-}C] = \{A, B, C\},$$

где $A : P \rightarrow Q$, $B : L \rightarrow M$, $C : X \rightarrow Y$.

3. УЛЬТРАМНОЖЕСТВЕННАЯ МОДЕЛЬ ГИПЕРТЕКСТА И УЛЬТРАТЕКСТА

Гипертекст удобно изображать графически. На рис. 1 изображен гипертекст, описывающий некоторую многоуровневую систему x . В системе x выделены подсистемы y_1, y_2, \dots, y_m , в которых в свою очередь выделены части z_1, z_2, \dots, z_n . Гипертекст состоит из отдельных текстов, соединенных между собой ссылками (связями). На рис. 1 обозначены через

$$(T-x), (T-y_1), (T-y_2), \dots, (T-y_m), (T-z_1), (T-z_2), \dots, (T-z_n)$$

отдельные тексты описания соответственно системы x в целом, подсистем y_1, y_2, \dots, y_m , а также частей этих подсистем z_1, z_2, \dots, z_n . Дуги на рис. 1 — это ссылки от целого к части. Рассмотрим подробнее простейший гипертекст

$$(T-x) \rightarrow (T-y)$$

с одной ссылкой C и двумя отдельными текстами $(T-x)$ и $(T-y)$, которые сами не являются гипертекстами, рис. 2.

Сделаем следующие допущения. Объект x принадлежит некоторому опорному множеству X , объект y принадлежит некоторому опорному множеству Y . Ссылка C в гипертексте определяется оператором, который будем называть опорным

$$C : X \rightarrow Y.$$

В частности $y = C(x)$.

Текст $(T-x)$ является элементом соответствующего ультрамножества $[\text{ULTR-}X] = P \times L \times X$, т.е. $(T-x) = (p)l(x)$. Здесь P — решетка достоверностей, L — решетка понятий (свойств объектов) опорного множества X . Аналогично текст $(T-y)$ является элементом соответствующего ультрамножества $[\text{ULTR-}Y] = Q \times M \times Y$, т.е. $(T-y) = (q)m(y)$. Тем самым построена ультрамножественная модель гипертекста.

Реализуя идею ультратекста, оснастим опорный оператор знаниями до ультраоператора

$$[\text{ULTR-}C] : [\text{ULTR-}X] \rightarrow [\text{ULTR-}Y].$$

В результате получим оперативную экспертную систему, соответствующую заданному гипертексту, рис. 3. ОЭС для заданного гипертекста можно немного усложнить, если рассмотреть обратные ссылки. В этом случае кроме прямого опорного оператора надо рассмотреть обратный

$$C\text{-BACK} : Y \rightarrow X$$

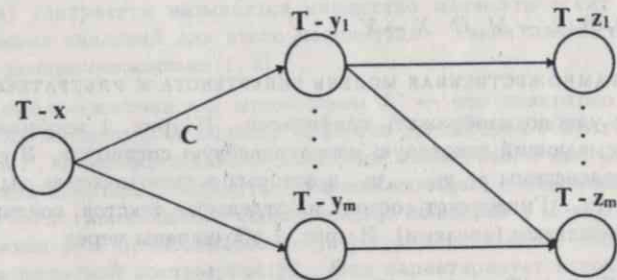


Рис. 1. Гипертекст.

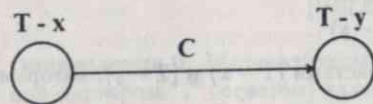


Рис. 2. Оператор.

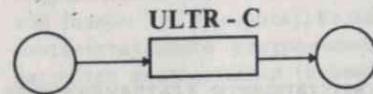
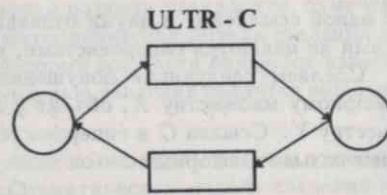


Рис. 3. Ультраоператор.



ULTR-C-BACK
Рис. 4.

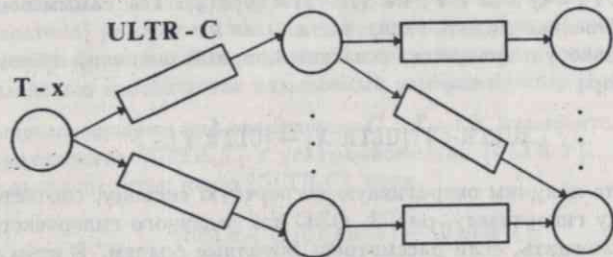


Рис. 5. Ультрасетевая экспертная система.

и оснастить его знаниями до ультраоператора

$$[ULTR-C-BACK]: [ULTR-Y] \rightarrow [ULTR-X].$$

В результате получим новую оперативную экспертную систему, соответствующую заданному гипертексту, рис. 4. Теперь введем заданный гипертекст в оперативную экспертную систему. В результате обработки системой гипертекста будет сгенерирован новый гипертекст

$$(T-x) \& [ULTR-C-BACK](T-y) \rightarrow (T-y) \& [ULTR-C](T-x),$$

который называется ультратекстом. Ультратекст имеет тот же граф из отдельных текстов, те же ссылки. Отличие его от заданного гипертекста состоит в том, что составляющие тексты ультратекста дополнены сведениями, полученными соответствующими ультраоператорами.

Аналогично строится ОЭС по любому гипертексту. Например, для гипертекста на рис. 1 ОЭС является система, изображенная на рис. 5.

Пример. Пусть имеется простейший гипертекст, рис. 2, в котором x - автомобиль, y - кузов этого автомобиля, $(T-x)$ - текст описания автомобиля в целом, $(T-y)$ - текст описания кузова этого автомобиля, C - ссылка от целого к части. На естественном языке ссылке $y = C(x)$ соответствует выражение "кузов автомобиля x ", обратной ссылке $x = C-BACK(y)$ соответствует выражение "автомобиль, кузов которого y ". Оснастим прямую ссылку знаниями и построим ультраоператор $[ULTR-C]$, в котором представим знания некоторого эксперта. Пусть среди правил в $[ULTR-C]$ есть такое: "Если возраст автомобиля более 5-ти лет и этот автомобиль эксплуатировался круглый год, то кузов этого автомобиля наверняка прогнил. Особенно дно, крылья, пороги". Правила в $[ULTR-C]$ могут быть разные. Это зависит от эксперта. Теперь при обработке гипертекста оперативной экспертной системой будут получены дополнительные сведения о коррозии кузова и добавлены к тексту $(T-y)$. Тем самым будет получен новый текст

$$(T-y) \& [ULTR-C](T-x)$$

о кузове этого автомобиля. Аналогичная ситуация с обратной ссылкой. Пусть среди правил ультраоператора $[ULTR-C-BACK]$ есть такое "Если бампер помят, то наверняка автомобиль попал в аварию". Теперь при обработке гипертекста оперативной экспертной системой будут получены дополнительные сведения об автомобиле, которые добавляются к имеющимся. Тем самым будет получен новый текст

$$(T-x) \& [ULTR-C-BACK](T-y)$$

об автомобиле в целом. Окончательно ОЭС генерирует новый гипертекст, который и называется ультратекстом.

СПИСОК ЛИТЕРАТУРЫ

1. Чечкин А.В. Математическая информатика. — М.: Наука, 1991г.
2. Чечкин А.В. Ультрамедиа — новое направление в искусственном интеллекте. (Настоящий номер журнала).
3. Conclin J. Hypertext: an introduction and overview // Computer, september, 1987. — N. 9. — V. 20. — P. 17-41.

К вопросу о понятии "цена информации"

Г.П. Шанкин

Рассматривается вопрос о цене (ценности) информации для конкретного пользователя. При этом ценность есть "прибыль", получаемая пользователем при реализации информации (с учетом наличия у него априорных знаний об интересующем его объекте). В конкретной семантической модели информации как сведений, заключенных в сообщении, исследуются такие свойства информации, как ее новизна, объем, достоверность, и т.д.

ВВЕДЕНИЕ

Вопрос о ценности (цене) информации¹ возникает при решении многих прикладных задач. В качестве примера укажем на две проблемы, при исследовании которых естественным образом возникает необходимость формализованной оценки информации.

Информационный обмен нередко принимает "рыночные" формы. Взгляд на информацию как товар (с присущими ему свойствами) получил достаточно широкую поддержку. В этом смысле можно сослаться, например, на работы [1]–[4]. Таким образом, за информацию необходимо платить, и возникает вопрос о разумном определении ее цены. Вторая проблема связана с обеспечением информационной безопасности. Здесь возникает вопрос о целесообразности вложения средств в защиту, т.е., по сути дела, об оптимизации этого вложения по критерию "цена защиты — цена защищаемой информации". В этой связи можно указать работы [5] и [6].

Существует немало работ, в которых предпринята попытка формализации понятия "цена информации", представления этой цены в виде функции от некоторых параметров, характеризующих ценностные (прагматические, аксиологические) свойства информации (см., например, [7]–[13]). Тем не менее, проблема формализованной оценки информации требует дальнейшего исследования. В данной статье автор высказывает некоторые собственные соображения по этому поводу.

Перейдем к изложению содержания статьи.

1. ИСПОЛЬЗУЕМ СЛЕДУЮЩИЕ ПРЕДПОЛОЖЕНИЯ И ОБОЗНАЧЕНИЯ

Множество текстов (сообщений), каждый из которых содержит информацию, подлежащую оцениванию, обозначается через X . Это множество

¹ Понятие цены и ценности информации в работе отождествляются.