УДК 371.263+004.82
ББК 74.00

ОЦЕНКА СЛОЖНОСТИ УЧЕБНОГО ТЕКСТА НА ОСНОВЕ СЕМАНТИЧЕСКОЙ СЕТИ

Наумов И. С. [1]
(*Учреждение Российской академии наук
Институт проблем управления РАН, Москва*)

Предложено решение задачи оценки сложности учебного текста на основе семантической сети. Для подсчета объема знаний в семантической сети разработан математический аппарат, базирующийся на определения семантических расстояний между понятиями. Показано, что объем знаний, содержащихся в семантической сети, является мерой на множестве семантических сетей, а введенное расстояние превращает это множество в метрическое пространство.

Ключевые слова: сложность текста, семантические сети, семантичское расстояние, знания, измерение знаний

1. Введение

С появлением большого объема текстовой информации возникло множество задач связанных с извлечением и обработкой данных из текстов. К актуальным задачам, в первую очередь, стоит отнести задачи поддающиеся максимальной автоматизации и не требующие привлечения экспертов предметной области. Среди таких задач стоит отметить задачу семантического поиска по тексту, извлечение знаний из текстов, генерация текстов. Решение перечисленных задач требует существенного развития методов и методик как в области лингвистики для обеспечения качественного синтаксического и семантического

[1] *Игорь Савельевич Наумов, инженер* (*naigsa@gmail.com*).

анализа, так и в области психологии для изучения механизмов получения, хранения и извлечения знаний человека.

Ведущую роль среди огромного объема текстовой информации стоит отвести учебным текстам. Учебный текст является основным средством получения знаний.

В отличие от остальных типов текстов учебный текст обладает структурой поддающейся наибольшей формализации: большая часть предложений обладает предикативной структурой, предложения согласованы таким образом, что выстраивают цепочку рассуждений, текст имеет однозначное толкование. Это позволяет применять автоматические синтаксические анализаторы для получения формальной структуры предложений в тексте.

Одной из важных задач в области анализа текстов является оценка сложности учебных текстов.

2. Сложность текста

Характеристики текста, влияющие на понимание и усвоение текста, называются сложностью текста. «Понимание текста есть осознание связей между элементами текста и объектами реального мира, которые обозначают эти элементы текста» [10].

В психологии понимание текста принято делить на три уровня: понимание слов, понимание предложений, понимание смысла одного или нескольких абзацев. Уровень понимания слов требует установления соотношения между словами и понятиями, которые эти слова обозначают в рассматриваемой предметной области. Понимание предложений требует установление связей между понятиями. На последнем уровне понимания происходит осознание основной идеи текста.

В [10] выделяются следующие компоненты сложности текста: информативность текста, сложность предложений, ясность структуры текста и абстрактность изложения. Ясность структуры и абстрактность текста можно установить только путем привлечения экспертов предметной области. Сложность предложений можно установить автоматически путем синтаксического анализа.

Информативность текста выражается тремя видами информации: содержательно-фактуальная (сведения о фактах, явлениях, событиях, действиях, лицах), содержательно концептуальная (авторское понимание обсуждаемых тезисов текста), содержательно-подтекстовый (скрытый смысл текста не выраженный синтаксическими структурами явным образом) [4].

Единственным видом информации, который позволяет извлекать информацию автоматически, является содержательно-фактуальная информация. В дальнейшем под сложностью текста будем понимать именно эту составляющую сложности текста.

Отражаясь в сознании читателя, содержательно-фактуальная информация приобретает статус смыслового представления текста. Для определения смысловой организации текста используется концептуальный анализ.

3. Концептуальный анализ текстов

Для выделения смыслового представления текста используется процедура концептуального анализа. Она позволяет выделить в тексте понятия и семантические связи между понятиями.

В [18] приведена классификация разработанных на сегодня концептуальных моделей текста. Первая модель основана на элементарной единице смысла – понятие. На базе разработанного ранее словаря предметной области в тексте выделяются понятия. После на основе подсчета частот понятий строится частотный словарь, который является концептуальной моделью текста.

В основе этой модели лежит положение о том, что структура текстов разных областей знаний неодинакова. Выделяется две лингвистические закономерности: в любом тексте используется лишь часть понятий предметной области, понятия предметной области используются неравномерно (одни понятия используются часто, другие – реже).

Такой класс моделей получил наибольшее распространение в методах сравнения текстов. Одним из популярных методов является метод латентно-семантического анализа (LSA).

Во второй модели за единицу смысла взято предложение. Основой предложения является предикат. Предикат выражает некоторое суждение, о котором говорится в предложении. Еще в дидактических исследованиях в качестве единицы информации использовалось суждение [16]. Подобные модели используются в исчислении предикатов.

В третьей модели за единицу смысла взято сверхфразовое единство. Сверхфразовое единство формируется посредством связи двух и более предложений связанных одной темой. Сверхфразовые единства используются в методах определения тематики текста [6].

Процедура концептуального анализа текста позволяет извлечь из текста декларативные знания. Одной из удобных и наглядных моделей представления знаний является семантическая сеть.

4. Семантическая сеть

Сематическая сеть позволяет определить объекты предметной области (вершины) и отношения между ними (отношения).

Пусть заданы две семантические сети: сеть текста S и сеть простого предложения S'. Сеть текста S зададим в виде упорядоченного множества из трех элементов:

(1) $S = (N, E, P)$,

где $N = \{n_i \mid i = \overline{1, Q'}\}$ – множество узлов сети с числом элементов Q, $E = \{(n_i, n_j, p_k) \mid n_i, n_j \in N; p_k \in P\}$ – множество ее дуг, P – множество двуместных предикатов. Дуги заданы упорядоченными множествами из трех элементов $(n_i, n_j, p_k) \in N \times N \times P$, где $n_i \in N$ – начальный узел, $n_j \in N$ – конечный узел, $p_k \in P$ – имя дуги, а $\times$ – знак операции декартового произведения множеств.

В свою очередь сеть предложения S' простая и состоит из двух узлов n_1, n_2 и одной дуги, помеченной именем некоторого предиката p:

$$S'' = (N', E', P'), \quad N' = \{n_1, n_2\}, \quad E' = \{(n_1, n_2, p)\},$$

Тогда объединением сетей $S=(N,E,P)$ и $S'=(N',E',P')$ будет сеть $S''=S\cup S'$ такая, что $S''=(N'',E'',P'')$ и $N''=N\cup N'$, $E''=E\cup E'$, $P''=P\cup P'$.

Аналогично можно определить пересечение сетей. Пересечением сетей $S'=(N',E',P')$ и $S''=(N'',E'',P'')$ будет сеть $S=S'\cap S''$ такая, что $S=(N,E,P)$ и $N=N'\cap N''$, $E=E'\cap E''$, $P=P'\cap P''$.

В свою очередь разностью сетей $S'=(N',E',P')$ и $S''=(N'',E'',P'')$ называется сеть $S=S'\setminus S''$ такая, что $S=(N,E,P)$ и $N=N'\setminus N''$, $E=E'\setminus E''$, $P=P'\setminus P''$.

Для упрощения семантические сети будем изображать в виде взвешенного ориентированного мультиграфа. В этом случае множество кратных дуг будем обозначать одной дугой с весом равному количеству кратных ей дуг. Так как исследуются не статистические свойства текста, а знания, этим текстом выражаемые, то при задании кратности дуг не будем учитывать частоту повторения однотипных отношений. В итоге имеем, что кратность дуги между двумя узлами сети равна числу различных типов отношений, связывающих соответствующие понятия.

Используя автоматические синтаксические анализаторы семантическая сеть текста может быть получена автоматически. В [13] описан метод основанный на предикативной структуре предложений. В основе метода лежит положение о том, что все предложения русского языка имеют предикативную структуру. Предикат описывает отношение объекта и субъекта. Таким образом, можно построить семантическую сеть текста, в которой вершинами являются объекты и субъекты предложений, выражающие понятия предметной области, а отношениями являются предикаты.

5. *Семантическое расстояние*

Два текста выражают близкие знания, если пересечение их семантических сетей соизмеримо с их объединением. В то же время два текста выражают разные знания, если пересечение семантических сетей мало по сравнению с их объединением.

Для количественной оценки близости текстологических знаний необходимо уметь определять расстояние между семантическими сетями и узлами одной семантической сети.

Большинство существующих методов определения семантического расстояния между понятиями опираются на определенные типы отношений, характерные для таксономий [13]. Выявление таких отношений в тексте является нетривиальной задачей, требующей, в конечном итоге, привлечения экспертов предметной области [14].

Для рассматриваемого класса задач семантическое расстояние между понятиями определим иным образом. Пусть задана семантическая сеть S. Зафиксируем два произвольных ее узла n_i и n_j. Найдем $R(n_i, n_j)$ – множество путей без циклов (цепей) длины не более чем M, ведущих от узла n_i к узлу n_j. Тогда семантическое расстояние L между узлами n_i и n_j может быть вычислено по формуле:

$$(2) \quad L(n_i, n_j) = \sum_{r \in R(n_i, n_j)} \frac{\min(w_1^r, w_2^r, \ldots, w_{d(r)}^r)}{d(r)},$$

где $d(r)$ – длина пути r, $d(r) \le M$; M – глубина связи; w_i^r – вес дуги i пути r, $i = \overline{1, d(r)}$, min – функция, возвращающая минимальное значение ее аргументов.

Из формулы (2) следует, что два узла отдалены друг от друга, если между ними имеется много путей (понятия слабо связаны). Отдаленность двух узлов тем больше, чем больше веса соединяющих их дуг (более вариативными являются связи между понятиями). Однако если в пути встречается дуга с небольшим весом, то этот путь вносит меньший вклад в удаленность узлов друг от друга. Но не все пути учитываются при подсчете расстояния между узлами: исключаются те пути, длина которых больше заданной глубины связи (трудно установить связь между понятиями, так как это требует использования большого числа предложений).

Следует обратить внимание на то, что семантическое расстояние (2) не является метрикой, так как для него не выполня-

ются аксиомы тождества, симметрии и неравенства треугольника. Это связано с тем, что понятие может иметь отношение к самому себе, связи между двумя понятиями по определению не симметричные и близость двух понятий зависит не только от непосредственно связывающих их предикатов, но и от предикатов, задающих косвенную связь через другие понятия.

Два понятия, соединенные длинным путем, признаются слабо связанными. Поэтому, при расчете семантических расстояний задается глубина учитываемых связей M.

Известно, что межсловесную связь можно осознать, если соответствующие слова хранятся в кратковременной памяти [10]. Существует множество работ посвященных изучению кратковременной памяти [3, 11, 12, 15]. Согласно этим работам длина кратковременной памяти равна 6-8 несвязанным единицам. Под единицами понимаются слоги, слова, предикаты, предложения. Например, сложно запомнить слово состоящее из боле чем 8 слогов, сложно понять предложение с первого прочтения состоящее более чем из 8 слов, сложно проследить рассуждение состоящее более чем из 8 предикатов.

Поэтому, для многих практических применений M можно выбирать из диапазона от двух до семи. Стандартная интерпретация M – число суждений, которыми одновременно может оперировать обучающийся, или максимальное число суждений, встречающихся в его умозаключениях.

Таким образом, единицей измерения семантических расстояний является семантическое отношение, описывающее отношение двух понятий.

В случае если два узла сети не связаны ни одним путем, то вычисление семантического расстояния между ними дает величину, равную нулю. Нулевое семантическое расстояние обозначает отсутствие связи между соответствующими понятиями и утверждает их содержательную неразличимость. С другой стороны, чем больше величина семантического расстояния между узлами сети, тем более отдаленными являются соответствующие понятия по содержанию.

Показательно значение семантического расстояния $L(n,n)$ между одним и тем же узлом n. Если $L(n,n)$ равно нулю, то соответствующее понятие может быть признано простым. Если величина $L(n,n)$ большая, то соответствующее понятие является сложным и может быть признано как не раскрытое по содержанию.

6. *Измерение знаний*

Помимо семантического расстояния между узлами сети для оценки сложности учебных текстов требуется вычисление семантического расстояния между семантическими сетями и определение объема знаний в них содержащихся.

Измерение объемов знаний до сих пор осуществляется методами, основанными на экспертных оценках результатов учебной работы обучающихся (субъективные методы) и на тестировании обучающихся (объективные методы). Теоретический фундамент этих методов заложен в современной теории педагогических измерений [1], где процесс обучения рассматривается как постоянное преодоление обучающимся грани между доступной областью знаний (уровнем актуального развития) и потенциально доступной (зоной ближайшего развития). Задача педагогов состоит в том, чтобы подобрать трудные, но посильные задания, способствующие выявлению уровня актуального развития [7].

Известен также подход, согласно которому измерение знаний осуществляется на основе измерения емкости понятий, где под емкостью понятия понимается число связей этого понятия с другими понятиями, а сама единичная связь выступает в качестве единицы измерения [9]. В этом случае измерение объема знаний в тексте, теме, учебной дисциплине сводится к выявлению понятий предметной области и подсчету числа связей между ними экспертными методами.

Для определения близости двух семантических сетей используется поиск гомоморфизмов, преобразующего одну сеть в другую. Однако нахождение гомоморфизма позволяет опреде-

лить только качественную «похожесть» сетей и не позволяет измерить объемы знаний, содержащихся в этих сетях.

Очевидно, что перечисленные методы непригодны для определения объема знаний в семантической сети текста, полученной на основе синтаксического анализа, так как в одном случае требуется привлечение экспертов (экспертная оценка, тестирование, подсчет емкости понятий), а в другом случае – отсутствует эффективно вычисляемое расстояние между семантическими сетями (поиск гомоморфизмов).

Под объемом знаний, содержащихся в семантической сети $S = (N, E, P)$, будем понимать величину, вычисляемую по следующей формуле:

(3) $K(S) = \sum_{n_i, n_j \in N} L(n_i, n_j)$,

где $K(S)$ – объем знаний в семантической сети S, а $L(n_i, n_j)$ – семантическое расстояние между узлами n_i и n_j, вычисляемое по формуле (2).

Формула (3) утверждает, что объем знаний в сети S есть сумма семантических расстояний между всеми парами ее узлов.

Как и у семантического расстояния, единицей измерения объема знаний является семантическое отношение.

Теорема 1 – Объем знаний (3) является аддитивной мерой на множестве семантических сетей.

Для доказательства теоремы 1 сначала покажем, что мера пустой сети равна нулю. Действительно, если семантическая сеть S пуста, $S = (\varnothing, \varnothing, \varnothing)$, то из формул (3) и (2) непосредственно следует $K(S) = 0$. Также из формул (3) и (2) следует утверждение о том, что мера объединения двух сетей S_1 и S_2 таких, что $S_1 \cap S_2 = (\varnothing, \varnothing, \varnothing)$, равна сумме их мер: $K(S_1 \cup S_2) = K(S_1) + K(S_2)$. ♦

Таким образом, в отличие от других известных подходов, формула (3) позволяет объективно измерить объем знаний, содержащийся в произвольном тексте.

Известны также несколько подходов к определению расстояний между графами. Это использование высоты ориентиро-

ванного графа, которая равна наибольшей длине пути от корня к листу в ярусно-параллельной форме его представления [17]. Также используется расстояние, получаемое на основе вычисления диаметра графа – максимального числа ребер, связывающих две его вершины [5]. Известно также расстояние между графами, получаемое путем вычисления реберной плотности – числовой величины, характеризующей близость графа к полносвязному [8].

Очевидно, что перечисленные подходы непригодны для определения расстояний между семантическими сетями. Семантическое расстояние между сетями S_1 и S_2 определим как объем знаний, содержащийся в симметрической разности этих сетей:

(4) $D(S_1, S_2) = K(S_1 \setminus S_2 \cup S_2 \setminus S_1)$.

Теорема 2 – Семантическое расстояние (4) является метрикой на множестве семантических сетей.

Для доказательства теоремы 2 достаточно показать, что на множестве семантических сетей удовлетворяются аксиомы тождества, симметрии и треугольника. Пусть S_1 и S_2 – семантические сети. Если $S_1 = S_2$, то из (4) следует $D(S_1, S_2) = 0$. Теперь пусть $D(S_1, S_2) = 0$. Тогда из (4) следует, что $S_1 = S_2$. В итоге $D(S_1, S_2) = 0$ тогда и только тогда, когда $S_1 = S_2$. Аксиома тождества доказана. Аксиома симметрии также непосредственно следует из (4): $D(S_1, S_2) = D(S_2, S_1)$. В свою очередь аксиома треугольника следует из формулы (4) и теоремы 1:

$K(S_1 \setminus S_2 \cup S_2 \setminus S_1) + K(S_2 \setminus S_3 \cup S_3 \setminus S_2) \geq K(S_1 \setminus S_3 \cup S_3 \setminus S_1)$. ♦

Таким образом, множество семантических сетей текстов является метрическим пространством, а семантическое расстояние между двумя сетями равно суммарному объему знаний, в них содержащихся.

7. Оценка знаний

Оценка знаний предполагает сопоставление имеющихся знаний с эталонными. В нашем случае эталонными знаниями

являются знания, выраженные текстом описания предметной области, а оцениваемые знания получены из произвольного текста, относящегося к этой же предметной области.

Рассмотрим предлагаемый в настоящей статье подход для оценки объемов знаний, а также сложности учебного текста. Пусть имеются две семантические сети:

– S – семантическая сеть предметной области, полученная из одного или более текстов;

– T – семантическая сеть оцениваемого текста;

Результат сопоставления сети предметной области S и сети текущей задачи T определяет сложность текста.

Учитывая специфику учебного текста можно показать, что сети S и T согласованы, т.е. $T \subseteq S$.

В итоге имеем формулу, позволяющие вычислить сложность учебного текста T:

(5) $C(T) = K(T \cap S)$, $c(T) = K(T \cap S) / K(S)$;

Сложность C текста T определяется по формуле (5) и равна объему знаний из семантической сети предметной области S, содержащейся в семантической сети текста T. Относительная сложность текста c – это доля знаний предметной области, содержащихся в тексте T.

8. Вычислительный эксперимент

Для проведения вычислительного эксперимента разработана программа (рисунок 1), которая позволяет загружать файлы с готовой семантической сетью в формате GraphML. Для работы с семантическими сетями в программе реализованы следующие инструменты:

– добавление, удаление, объединение семантических сетей;

– перемещение семантической сети в разные группы (тексты предметной области, оцениваемые тексты);

– просмотр статистических данных семантических сетей (частота слов и отношений, части речи);

– просмотр семантических сетей в графическом режиме.

На примере главы «Информация и информационные процессы» из учебника по информатике [2] построены семантические сети по каждому параграфу главы.

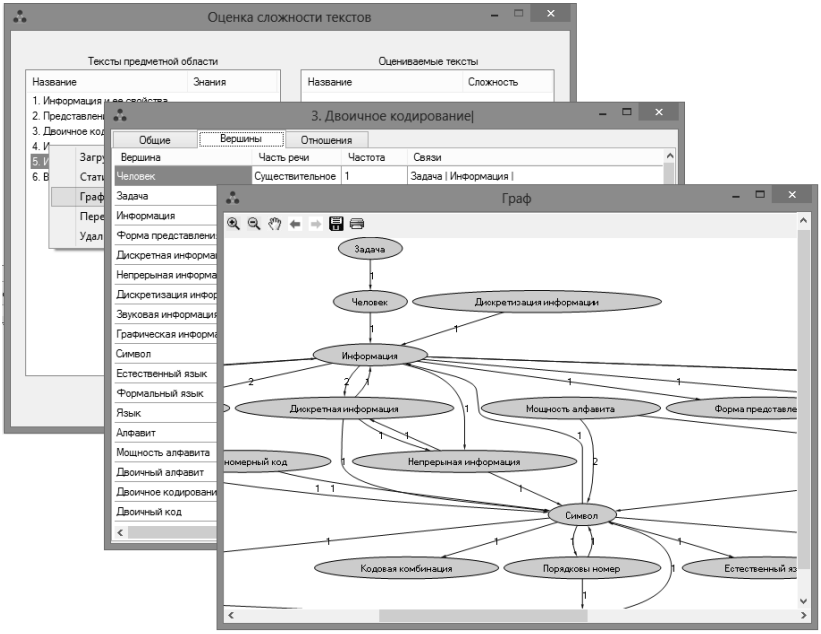

Рис. 1. Интерфейс программы

Используя формулу 6 для подсчета объема знаний, можно определить относительные объем знаний каждого параграфа (рисунок 2). Из диаграммы представленной на рисунке можно сделать вывод, что наибольший объем знаний несет в себе первый параграф. Это характерно для всех учебных пособий, так как в начале описания любой предметной области вводятся все базовые понятия, необходимые для понимания предшествующих разделов учебного материала.

Используя формулу 7 расчета сложности семантической сети, можно вычислить объем имеющихся знаний текущего параграфа, которые уже были получены из предыдущих параграфов. Таким образом, объем новых знаний текущего параграфа можно

вычислить как разность знаний текущего параграфа и сложности текущего параграфа относительно всех знаний предшествующих параграфов. На рисунке 3 изображена диаграмма объема новых знаний в каждом из параграфов.

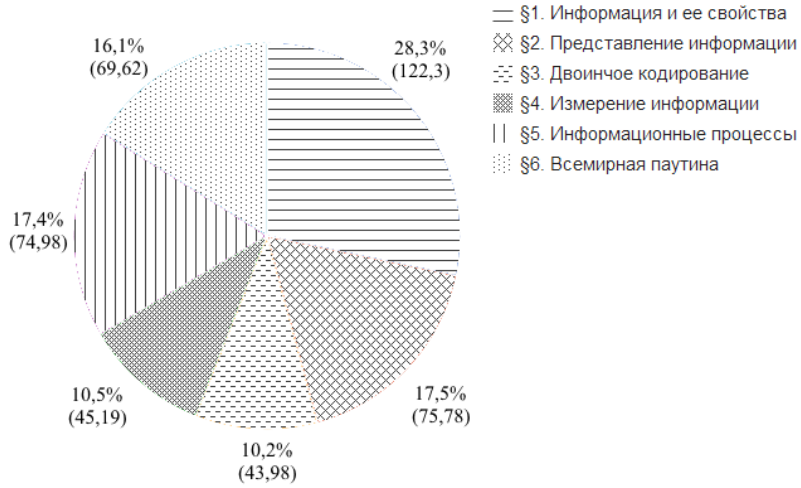

Рис. 2. Объем новых знаний главы по параграфам

Рис. 3. Объем новых знаний в параграфах

Из анализа полученных данных следует, что все параграфы данной главы обладаю высокой информативностью (более 70% знаний являются новыми). Также, все параграфы имеют относительно слабую связь с предыдущими параграфами (повторяется не более 10% знаний). Подобные особенности текста характерны именно для учебных текстов описания предметной области.

Таким образом, используя предложенный метод оценки сложности текста и метод подсчета объема знаний, можно выделить некоторые характеристики текста: объем знаний, распределение знаний по тексту, контекстная связь различных частей текста. Из этих данных, в свою очередь, можно сделать выводы о трудности восприятия текста.

9. Заключение

В данной работе было предложено решение задачи оценки сложности текстов и задачи измерения знаний, содержащихся в тексте.

Для оценки объемов знаний, содержащихся в семантических сетях, разработан математический аппарат, базирующийся на определения семантических расстояний между понятиями в семантической сети.

Показано, что объем знаний, содержащихся в семантической сети, является мерой на множестве семантических сетей, а введенное расстояние между семантическими сетями превращает множество семантических сетей в метрическое пространство.

В отличие от других существующих методов определения объемов знаний, основанных на использовании онтологий и тезаурусов, разработанный метод отличается универсальностью применения, так как не привязан к конкретной предметной области. Основное условие применимости метода – предоставление описаний предметной области в виде множества семантических сетей построенных по тексту.

Как и в теории информации Шеннона-Хартли [20, 21], так и в представленном методе измерения объемов знаний выполнено абстрагирование от психической природы изучаемых явлений и найден такой материальный объект, по характеристикам кото-

рого можно судить об интенсивности моделируемых психических процессов.

Так в теории информации – это вероятность сообщения, или частота его предъявления испытуемому: чем более часто появляется некоторое сообщение, тем оно менее «неожиданно» для испытуемого и, как следствие этого, воспринимается им как содержащее меньший объем информации. Для учета особенностей восприятия сообщений объем информации в сообщении определен как отрицательный логарифм от его вероятности.

В разработанном методе измерения объемов знаний – это текст, содержащий некоторые знания. Для учета активной природы знаний глубина прослеживаемых связей при подсчете объема знаний ограничена способностью человека устанавливать мысленные связи между понятиями путем выполнения умозаключений с определенным числом исходных суждений в них.

Литература

1. АВАНЕСОВ В.С. *Знания как предмет педагогического измерения* // Педагогические измерения. – 2005. – № 3. – С. 43-52.
2. БОСОВА Л.Л., БОСОВА А.Ю. *Информатика и ИКТ: учебник для 8 класса*. – 2-е изд., испр. – М.: БИНОМ. Лаборатория знаний. 2012. – 220 с.
3. ВЕЛИЧКОВСКИЙ Б.М. *Зрительная память и модели переработки информации человеком* // Вопросы психологии. – М. 1977. С 49-61.
4. ГАЛЬПЕРИН И.Р. *Текст как объект лингвистического исследования.* – М.: Изд-во «КомКнига», 1981. – 144 с.
5. ЕВСТИГНЕЕВ В.А. *Применение теории графов в программировании.* – М.: Наука, 1985. – 332 с.
6. ЕРМАКОВ А.Е. *Тематический анализ текста с выявлением сверхфразовой структуры* // Информационные технологии. 2000. – С. 37-40.

7. ЕФРЕМОВА Н.Ф. *Тестовый контроль в образовании: Учебное пособие.* – М.: Университетская книги, 2007. – 540 с.
8. КАРПЕНКО А.П., СОКОЛОВ Н.К. *Меры сложности семантической сети в обучающей системе* // Вестник МГТУ им. Н.Э. Баумана, серия «Приборостроение». – 2009. – №1 (74). – С. 50-66.
9. КАРПЕНКО М.П. *Проблема измерения знаний и образовательные технологии* // Журнал практического психолога. – 1997. – № 4. – С. 74-79.
10. МИКК Я.А. *Оптимизация сложности учебного текста*: *В помощь авторам и редакторам.* – М: Просвещение, 1981. – 33 с.
11. МИКК Я.А. *Понятность учебного текста и связи в нем // Советская педагогика и школа*. Вып. 2. – М: Тарту. 1970. С 5-72.
12. МИЛЛЕР Дж.А. *Магическое число семь плюс или минус два. О некоторых пределах нашей способности перерабатывать информацию* // Инженерная психология. – М. 1964. С. 192-225.
13. НАУМОВ И. С., ВЫХОВАНЕЦ В. С. *Оценка трудности и сложности учебных задач на основе синтаксического анализа текстов* / Управление большими системами. Выпуск 48. – М.: ИПУ РАН, 2014. С.97-131.
14. НАЙХАНОВА Л.В. *Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования*. – Улан-Удэ: Изд-во БНЦ СО РАН, 2008. – 237 с.
15. НЕВЕЛЬСКИЙ П.Б. *Исследование объема кратковременной и долговременной памяти* // Проблемы инженерной психологии. – М. 1967. С. 128-133.
16. ПОДЛАСЫЙ И.П. *К вопросу о надежности информационно-смысловых элементов текста (ИСЭТ)* // Новые исследования в педагогических науках. – М.:1972. С. 64-68.
17. ФЕДОТОВ И.Е. *Некоторые приемы параллельного программирования: Учебное пособие.* – М.: Изд-во МГИРЭА, 2008. – 188 с.

18. ХОРОШИЛОВ А.А. *Методы автоматического установления смысловой близости документов на основе их концептуального анализа.* – М.: ЦИТиС, 2013. – 8 с.
19. DUBAY W. *The Principles of Readability* // Impact Information. – Costa Mesa, California, 2004. – 73 p.
20. HARTLEY R.V.L. *Transmission of Information* // Bell System Technical Journal. – July 1928. – PP. 535–563.
21. SHANNON C.E. *A Mathematical Theory of Communication* // Bell System Technical Journal. – 1948. – V. 27. – PP. 379-423, 623-656.