

4. Gonzales R. Digital Image Processing. Boston : Wesley, 2003.
5. Duda R., Hart P. Pattern Classification. New York : Wiley, 2001.
6. Левтин К. Э. Адаптивная система обнаружения задымления на открытых пространствах в видеоматериалах // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций : материалы 16-й Междунар. науч.-техн. конф. / Рязан. гос. радиотехн. ун-т. Рязань, 2010. С. 121–122.
7. Левтин К. Э. Детектирование дыма в видеопотоках на основе вейвлет-преобразования // Решетневские чтения : материалы XIII Междунар. науч. конф. / Сиб. гос. аэрокосмич. ун-т. В 2 т. Ч. 2. Красноярск, 2009. С. 514–515.
8. Celik T., Ozkaramanl H., Demirel H. Fire and Smoke Detection Without Sensors: Image Processing Based Approach // Proc. of the Europ. Signal. Processing Conf. Poznan, 2007. P. 1794–1798.
9. Russel S., Norvig P. Artificial Intelligence: a Modern Approach. 3rd ed. Upper Saddle River, N. J. : Prentice Hall, 2010.
10. Celik T., Ozkaramanl H. Smoke and Fire Detection Using Video Sequence // Europ. Signal Processing Conf. Poznan, 2008. P. 197–200.
11. Cormen T. Introduction to Algorithms. Cambridge, Mass., 2006.
12. Левтин К. Э. Применение вейвлет-анализа к детектированию дыма на открытых пространствах // Актуальные проблемы авиации и космонавтики : материалы VI Всерос. науч.-практ. конф. студентов, аспирантов и молодых специалистов / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2010. С. 359–360.
13. Левтин К. Э. Блочный-текстурный метод детектирования дыма в видеопоследовательностях // Теоретические и прикладные вопросы современных информационных технологий : материалы X Всерос. науч.-техн. конф. Улан-Удэ, 2009. С. 248–250.

К. Е. Levtin

VISUAL SMOKE DETECTION BASED ON SPATIAL-TEMPORAL ANALYSIS OF VIDEO SEQUENCES

A hybrid approach to visual smoke detection, based on spatial-temporal clustering of moving objects, is proposed. Visual smoke detection system, based on the algorithm of the hybrid approach, was developed and tested on a whole number of video materials. Conclusions and possible further investigations are offered in the end of this paper.

Keywords: visual smoke detection, spatial-temporal clustering.

© Левтин К. Э., 2012

УДК 519.682

Д. В. Личаргин

МЕТОДЫ ГЕНЕРАЦИИ ПРЕДЛОЖЕНИЙ ЕСТЕСТВЕННОГО ЯЗЫКА НА ОСНОВЕ ЛЕСА ДАННЫХ ЕСТЕСТВЕННОГО ЯЗЫКА

Рассматриваются модели и средства генерации осмысленных единиц, в частности предложений естественного языка, предназначенные для решения проблем компьютерной лингвистики. При этом на основе леса деревьев классификаций задаются иерархия наследующих друг друга информационных единиц с признаками классификаций более низкого уровня и семантическое понятийное пространство языка, соответствующее симметричному дереву, на каждом уровне и срезе леса естественного языка. На последнем уровне задается множество функций и деревьев функций, соответствующих осмысленным предложениям естественного языка, и другие гнездящиеся над основной классификацией данные.

Ключевые слова: генерация естественного языка, системный анализ, классификации слов языка.

На современном этапе актуальной является задача автоматизации систем письменного и устного перевода для различных языков, экспертных, поисковых систем и систем реферирования. Для решения данной задачи успешно используются многочисленные теории, концепции и программные системы, а работы в области семантики, дискретной математики, лингвистики и искусственного интеллекта дают надежду на преодоление в ближайшем будущем многих проблем

формализации естественного языка и прохождение теста Тьюринга во все более жестких для тестовых систем условиях.

Для генерации осмысленной речи в настоящее время используется широкий инструментарий как семантики, так и искусственного интеллекта в рамках понятийного аппарата и различных моделей математической семантики [1]. В частности, для анализа естественного языка традиционно применяются такие

модели и средства, как метод онтологий, метод лингвистической классификации, метод многомерного представления данных, OLAP-системы, реляционные базы данных, фреймы, порождающие грамматики (например, порождающие грамматики Монтегю), семантические сети, теория графов и метод резолюций, гибридные системы, а также лингвистические методы: компонентный анализ, валентностное представление слов языка, парадигматический метод, американский структурализм и др.

Однако трудности в создании приемлемых алгоритмов машинного перевода с одного естественного языка на другой [2] заставляют говорить о необходимости более глубокого исследования семантики языка. Остро стоит проблема формализации смысла языковых единиц и их единой классификации.

В данной статье приведено понятийное описание единиц естественного языка, позволяющее использовать критерии осмысленности фраз на естественном языке, описанные в работе [2], на основе порождаемого леса данных естественного языка как системы информационных единиц [3].

Разработанный автором словарь порождения высказываний является главной составляющей предлагаемой модели естественного языка как системы иерархий языковых единиц.

Порождение понятийного пространства единиц естественного языка. Рассмотрим семантическую классификацию слов и понятий естественного языка, сводимую к шестнадцати классам сем (семантических, смысловых атомов) языка и, далее, к четырем геносемам (элементарным частицам смысла), а также к понятию связи (кванта смысла), что может быть показано на основе понятийного аппарата семантических сетей.

Определение на основе кванта смысла представляет собой семантическую сеть, включающую в себя элементарный семантический квант, т. е. связь, дуги которой несут в себе семантику понятия тождества некоторых элементов объектов, т. е. связи объектов. Эти классификации приводятся в работе [2].

Под лесом в математике традиционно понимается граф, множество несвязанных деревьев. С ним можно ассоциировать отношения тождества и семантические отношения узлов деревьев леса над графом семантической сети. Симметричным деревом будем называть дерево с одним набором признаков классификации на каждом уровне дерева. Такое дерево однозначно соответствует многомерному пространству единиц естественного языка.

Множество деревьев леса единиц языка упорядочивается на основе четверки размерностей ориентированного леса лингвистического текста (рис. 1):

1) уровней (Levels) леса – $F[L(i)[\], _, _, _]$;

2) срезов (Sections) леса – $F[_, S(j)[\], _, _]$;

3) гнездящихся деревьев леса (Nested Trees), например иерархической структуры слова и других информационных единиц, – $F[_, _, N(l)[\], _]$;

4) срезов гнездящихся деревьев леса – $F[_, _, _, SN(m)[\]]$.

Модель языка, лес текста. Текст на естественном языке может быть представлен как система элементов и связей между ними. Лингвистические системы являются сложными для процесса моделирования. Тем не менее можно выделить ряд основных компонентов лингвистических систем различных уровней (ярусов) языка и срезов (аспектов) языка.

Текст состоит из иерархии уровней/ярусов [4; 5]:

– множества бессмысленных текстов – $F[L(21)[\], _, _, _]$;

– множества грамматически осмысленных текстов – $F[L(20)[\], _, _, _]$;

– множества семантически осмысленных текстов – $F[L(19)[\], _, _, _]$;

– множества всех существующих текстов – $F[L(18)[\], _, _, _]$;

– библиотеки – $F[L(17)[\], _, _, _]$;

– классификации текстов в каталоге библиотеки –

$F[L(16)[\], _, _, _]$;

– серии книг – $F[L(15)[\], _, _, _]$;

– набора томов – $F[L(14)[\], _, _, _]$;

– тома – $F[L(13)[\], _, _, _]$;

– глав – $F[L(12)[\], _, _, _]$;

– разделов/параграфов – $F[L(11)[\], _, _, _]$;

– абзацев – $F[L(10)[\], _, _, _]$;

– паре и цепочек предложений – $F[L(9)[\], _, _, _]$;

– сложных предложений – $F[L(8)[\], _, _, _]$;

– простых предложений – $F[L(7)[\], _, _, _]$;

– конструкций – $F[L(6)[\], _, _, _]$;

– синтагм – $F[L(5)[\], _, _, _]$;

– фразеологизмов – $F[L(4)[\], _, _, _]$;

– словоформ – $F[L(3)[\], _, _, _]$;

– морфем – $F[L(2)[\], _, _, _]$;

– букв – $F[L(1)[\], _, _, _]$;

– признаков буквы – $F[L(0)[\], _, _, _]$.

Текст естественного языка состоит из следующих срезов/аспектов:

– среза написания (цепочки букв – символов алфавита) – $F[_, S(0)[\], _, _]$;

– среза произношения (цепочки звуков) – $F[_, S(1)[\], _, _]$;

– грамматического среза (добавления грамматических конструкций и категорий) – $F[_, S(3)[\], _, _]$;

– семантического среза (шаблонов подстановок смысловых единиц языка) – $F[_, S(4)[\], _, _]$;

– текстологического среза (шаблонов заполнения относительно статической структуры текста) – $F[_, S(5)[\], _, _]$;

– среза актуального членения предложения (тема, рема, модальность, пояснение и др.) – $F[_, S(6)[\], _, _]$;

– стилистического среза (множеств особенностей всех предыдущих срезов в зависимости от ситуации и манеры речи) – $F[_, S(7)[\], _, _]$ и др.

Каждый уровень рассматривается как набор информационных единиц в пространстве возможных состояний, т. е. семантическом понятийном пространстве морфем, слов, текстов и т. д.

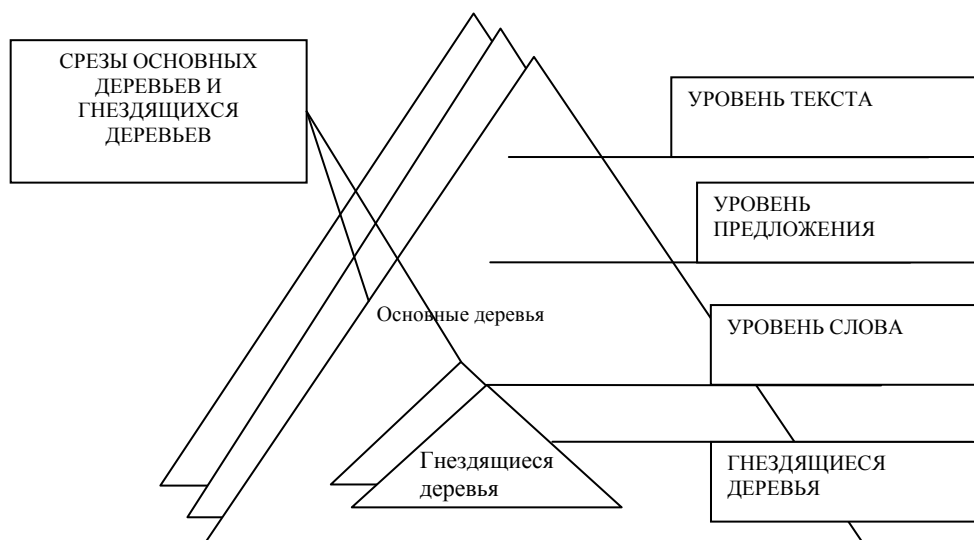


Рис. 1. Размерности леса данных естественного языка

Таблица 1

Распределенная структура слова в системе предложения естественного языка

Уровень	Активные валентности					Пассивные валентности			
	Валентность 1	Ядро слова	Валентность 2	Предлог	Валентность 3	Валентность A	Валентность B	Валентность C	Валентность D
Грамматика: члены предложения, части речи и категории	Субъект	Предикат, глагол	Объект, существительное	Связка, предлог	Реципиент (косвенный объект)	Обстоятельство места	Обстоятельство времени	Обстоятельство образа действия	Обстоятельство инструмента
Семантика	Существо	Отношение–существо–существо–предмет	Предмет	Отношение, ноль переходит в единицу	Существо	Место	Отношение–нечто (время)	Отношение–Отношение	Отношение–Существо–Объект
Написание	Любое	Give	Любое	to	Любое	Любое	Любое	Любое	Любое
Фонетика	Любое	[гИв]	Любое	[ту]	Любое	Любое	Любое	Любое	Любое
Стилистика, ...	Нейтральная	Обиходное понятие, обиходный стиль	Нейтральная	Нейтральная	Нейтральная	Нейтральная	Нейтральная	Нейтральная	Нейтральная

Уровни семантического среза естественного языка. Язык состоит из следующих типов семантических единиц языка разного масштаба (это семантический аспект языка):

- геносем $F[L(1)[\] , _ , _ , _]$ – самых мелких единиц языка, элементарных частиц смысла;
- сем (около двухсот) $F[L(2)[\] , _ , _ , _]$ – атомов смысла, составляющих структуру слов и понятий;
- понятий (произвольное количество) $F[L(3)[\] , _ , _ , _]$ – смысловых единиц, соответствующих словам;
- фактов $F[L(8)[\] , _ , _ , _]$ – элементов знания из различных областей и др.

Эта иерархия определяет вывод смысловых единиц обиходного естественного языка на основе элементарных составляющих (табл. 1).

Будем условно исходить из тезиса, что этих четырех видов единиц языка разного масштаба будет достаточно для редукционизма описания формализуемой информации языка, хотя очевидно, что любое подобное описание будет недостаточным. Так, например, с точки зрения абстрактной комбинаторики количество возможных классификаций намного превосходит количество элементов, над которыми эти классификации строятся.

Проекция семантической и формальной классификации друг на друга. Классификация любых единиц языка, в частности, слов и понятий, представляются в форме дерева, узлами которого являются единицы языка. Каждой семантической единице языка A может соответствовать некоторое количество n формальных единиц языка B_n , где $n \in \{0, 1, 2, 3, \dots\}$; где $A \subset F[_, S(4)[_, _, _]]$; $B \subset F[_, S(3)[_, _, _]]$, и, наоборот, каждой формальной единице языка B может соответствовать некоторое количество m семантических единиц языка A_m , где $m \in \{0, 1, 2, 3, \dots\}$. Таким образом, узлы дерева, описывающие семантическую классификацию понятий, проецируются на узлы дерева формальной классификации, описывающей словоформы. Если одному узлу формальной классификации соответствует несколько узлов семантической классификации, то последние будут называться различными лексико-семантическими вариантами (ЛСВ), т. е. различными значениями одного и того же слова, либо омонимами. Если одному узлу семантической классификации соответствует несколько узлов формальной классификации, то эти узлы будут называться синонимами.

Принципы построения классификации. Классификация понятий языка может быть представлена в форме дерева. Множество узлов дерева семантической классификации назовем понятийным пространством.

В дереве классификаций выделяются уровни. На одном уровне классификации может быть только один классификационный признак. Признаки разных уровней составляют ряд, или вектор, признаков заданной классификации. Каждому признаку классификации соответствует множество возможных значений данного признака. Все эти признаки являются сложными и состоят из конечного числа смысловых элементов. Каждый из этих элементов является узлом другой классификации меньшего масштаба. Таким образом, классификация понятий строится на основе элементов, классифицируемых в классификации сем.

Понятийное пространство предложений естественного языка. Рассмотрим многомерное грамматическое пространство единиц естественного языка: слов и предложений. Такое пространство слов позволяет генерировать грамматически, но не семантически осмысленные фразы естественного языка. Так, фраза «видеть я» является грамматически бессмысленной, фраза «я ем шляпу» – грамматически осмысленна, но семантически бессмысленна, а фраза «я ем грушу» – грамматически и семантически осмысленна. Для генерации семантически и грамматически осмысленных фраз естественного языка классификация предложений естественного языка может быть построена в виде многомерного пространства данных. Эта классификация является пересечением лексического и грамматического пространства слов – точек понятийного пространства. Грамматическое подмножество этой классификации имеет следующие координаты:

$F[L(\text{«слово»}), S(\text{«грамматика»}), N(\text{«распределенная структура слова»}), NS(\text{«аспект структуры слова»})]$.

Например, подмножество классификации может быть задано следующим вектором признаков с множествами возможных координат: $F[L(\text{«слово»}), S(\text{«грамматика»})$ [Части речи {«Артикль», «Прилагательное», «Существительное», «Глагол», ...}, Члены предложения {«Определитель», «Определение», «Подлежащее», «Сказуемое», ...}, Категории[Лица {«1-е», «2-е», «3-е», «Не определено»}, Аспект {«Неопределенный», «Продолженный», «Совершенный», «Совершенный продолженный», «Не определен», ...}], $N(\text{«слово»}), NS(\text{«грамматика»})$ («морфология») [предлог, артикль, именная группа, предлог] \supset {«in May», «care of», «the Internet», «caring of», «be the Internet», ...}.

Грамматические конструкции включаются в ячейки многомерного массива. Пересечение таких координат вектора, как, например, $F[L(\text{«слово»}), S(\text{«грамматика»}), N(\text{«распределенная структура слова»}), NS(\text{«грамматика»})$ [Части речи {«Артикль», «Прилагательное», «Существительное», «Глагол», ...}, Члены предложения {«Определитель», «Определение», «Подлежащее», «Сказуемое», ...}, Категории [Лица {«1-е», «2-е», «3-е», «Не определено»}, Аспект {«Неопределенный», «Продолженный», «Совершенный», «Совершенный продолженный», «Не определен», ...}], $N(\text{«распределенная структура слова»}), NS(\text{«грамматика»})$ («морфология») [«вспомогательный глагол», «форма глагола», «место предлога»] \supset «having + Глагол-ed», определяет соответствующую ячейку многомерного массива с грамматической конструкцией. Реляционные таблицы как подмножества этого многомерного массива представлены в лингвистике традиционными грамматическими парадигмами.

В свою очередь лексико-грамматическое подмножество классификации предложений естественного языка имеет следующие координаты: $F[L(\text{«предложение»}), S(\text{«лексика»})$ («грамматика») [Порядок слов и члены предложения {Субъект, Предикат, Объект}, Объекты по тематике изучения {идеи {науки, представления, чувства, ...}, предметы {одежда, еда, части тела, здания, транспорт, ...}, существа, ...}, Варианты подстановок слов в предложение {позитивное {обожать, любить, ...}, негативное {не любить, ненавидеть, ...}, ...}], $N(\text{«функции предложения над точками слов»}), NS(\text{«неопределенное»})]$.

Такое многомерное пространство включает в себя комбинаторно сочетающиеся группы слов. Например, группа слов {носить, одевать, снимать, гладить, шить, ...} относится к ячейке многомерного пространства $F[L(\text{«простое предложение»}), S(\text{«грамматика»}), N(\text{«слово как точка классификации»}), _] * F[L(\text{«простое предложение»}), S(\text{«семантика»}), N(\text{«слово как точка классификации»}), _] \supset S(\text{«грамматика»})$ [«отношение–существо–объект–предмет» / «одежда», «глагол», «предикат», «неопределенная форма»].

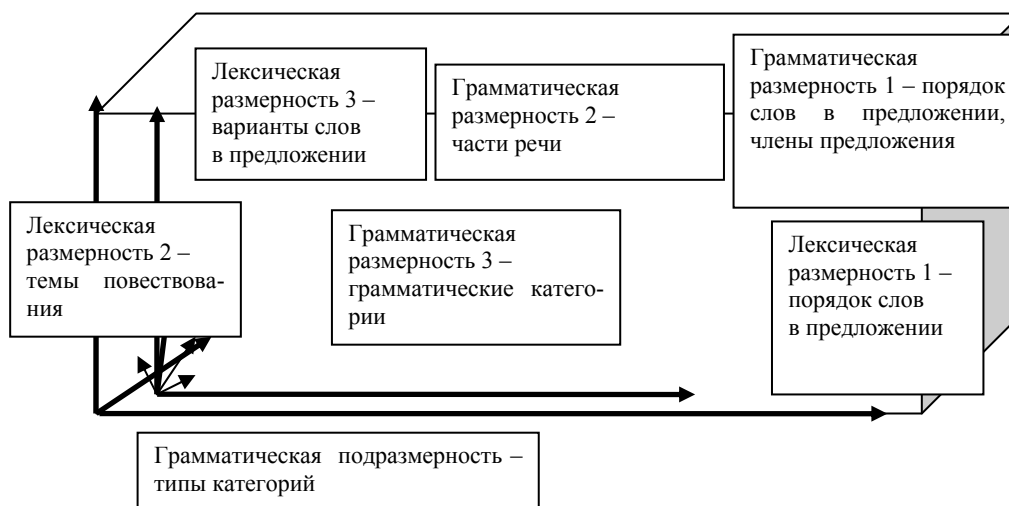


Рис. 2. Координаты многомерного лексико-грамматического подпространства леса данных естественного языка

Таблица 2

Принцип генерации осмысленных фраз естественного языка методом подстановки

the ... <i>Зэ ... этот ...</i>	of the ... <i>Ов Зэ ... этото ...</i>	is over <i>Из Оувэ закончится</i>	now <i>нАу сейчас</i>
series <i>сИэриз серия</i>	game <i>гЕйм игра</i>	is left <i>Из лЕфт осталось</i>	at the present moment <i>эт Зэ прЕэнт моумезнт в настоящем</i>
season <i>сИ:зн сезон</i>	tournament <i>тУэнэмэнт турнир</i>	starts <i>стА:тс начинается</i>	today <i>тудЕй сегодня</i>
cycle <i>сАйкл цикл</i>	Olimpic games <i>олиМпик гЕймз Олимпийские игры</i>	goes on <i>гОуз Он продолжается</i>	this week <i>ЗИс ВИ:к на этой неделе</i>

При этом группа слов {кофта, носки, куртка, майка, фартук, ...} относится к ячейке многомерного массива $F[L(\text{«слово»}), S(\text{«грамматика»})]$ («семантика») [«объект», «одежда», «существительное», «субъект», «единственное число»] $_, _$. Обе группы слов образуют синтагматические пары вида $F[L(\text{«синтагма»}), S(\text{«грамматика»})]$ («семантика») [«объект», «одежда», «существительное», «субъект», «единственное число»] + [«действие с объектом», «одежда», «глагол», «предикат», «неопределенная форма»], $N(\text{«функция двух аргументов»})$, $_, _$: «носить кофту», «гладить фартук», «шить носки», «снимать куртку» и т. п. Грамматический порядок слов получает в соответствие семантические групп слов, в результате чего данное пространство становится критерием семантической и грамматической осмысленности речи. Функции определенного вида и определенной геометрии над данными группами слов образуют осмысленные фразы с хорошей вероятностью. Фрагменты этих функций представляют собой предложения осмысленного естественного языка, а функции предложений соотносятся с гнездящимися деревьями уровня предложения.

Для решения проблемы нахождения критериев семантической осмысленности необходимо задать понятийное пространство единиц естественного языка, в частности в форме леса единиц естественного языка. Далее необходимо задать правила генерации функций осмысленных предложений и кластеров функций, соответствующих осмысленным текстам: $F[L(\text{«текст»}), S(\text{«грамматика»})]$ («семантика»), $N(\text{«кластеры функ-$

ций-предложений»), $_, _$. Функции определенного вида будут рассматриваться как осмысленные. Пространство, как и симметричные деревья единиц естественного языка, задается векторами классификации возможных единиц естественного языка. Так, в работе [2] дается вектор классификации семантики слов естественного языка и ставится задача нахождения векторов классификации для ряда базовых единиц языка при построении системы генерации осмысленных текстов на естественном языке.

Принцип генерации осмысленных предложений на естественном языке используется программами «Электронный словарь» и «Электронный разговорник» [2]. Пример подстановочной таблицы как среза многомерного понятийного пространства слов естественного языка приведен в табл. 2.

Таким образом, получена оригинальная модель пространства единиц естественного языка на основе леса векторно-упорядоченных деревьев данных естественного языка. В рамках этой модели ставится задача построения векторов подклассификаций: слов языка, синтагм, пар слов, повествований, диалогов, текстов и т. п. – для решения проблемы формализации естественного языка.

Библиографические ссылки

1. Automatic Expansion of Domain-Specific Lexicon by Term Categorization / H. Avancini, A. Lavelli, F. Sebastiani, R. Zanolli // ACM Transl. on Speech and Lang. Processing. 2006. Vol. 3, № 1. P. 1–30.

2. Личаргин Д. В. Методы и средства порождения семантических конструкций естественно-языкового интерфейса программных систем : дис. ... канд. техн. наук : 05.13.17. Красноярск, 2004.

3. Личаргин Д. В. Порождение дерева состояний на основе порождающих грамматик над де-

ревьями строк // Вестник СибГАУ. 2009. Вып. 4 (25). С. 33–37.

4. Никитин М. В. Лексическое значение слова. М. : Высш. шк., 1983.

5. Селиверстова О. Н. Контрастивная синтаксическая семантика. М. : Наука, 1990.

D. V. Lichargin

METHODS OF NATURAL LANGUAGE SENTENCES GENERATION BASED ON NATURAL LANGUAGE DATA FOREST

In the work the author describes models of sense bearing units generation, like natural language sentences, as means of solving the problems of computational linguistics. Therefore a hierarchy of informational units inherited by each other, based on the classification trees, is determined, with the features of the lower level classification. A semantic notional space of the natural language is defined, (corresponding to a symmetric tree) on each level and section of the forest of the natural language, where a set of functions and function trees are assigned, corresponding to the sense bearing sentences of the natural language, and other data, nested over the main classification, are assigned as well.

Keywords: natural language generation, system analysis, language words classification.

© Личаргин Д. В., 2012

УДК 519.62

В. А. Нестеров

КОНЕЧНО-ЭЛЕМЕНТНЫЙ АНАЛИЗ УСТОЙЧИВОСТИ БАЛОК, ПОДАТЛИВЫХ ПРИ ТРАНСВЕРСАЛЬНОМ СДВИГЕ

Рассматривается задача об устойчивости при осевом сжатии балки с низкой трансверсальной сдвиговой жесткостью. На основе теории пластин Рейсснера-Мидлина разработана конечно-элементная модель, в которой в качестве узловых неизвестных фигурируют углы трансверсального сдвига. Представлены решения, демонстрирующие актуальность разработанной модели. Исследуется устойчивость трехслойных и композитных балок при учете неклассических граничных условий.

Ключевые слова: балка, устойчивость, трансверсальный сдвиг, метод конечных элементов.

При рассмотрении задачи об устойчивости при осевом сжатии балок, податливых при трансверсальном сдвиге, решение выполняется с помощью метода конечных элементов (МКЭ). Математическая модель базируется на теории пластин Рейсснера-Мидлина. Для вывода основополагающих матриц МКЭ используется вариационный подход. Для этого в геометрически нелинейной постановке получено выражение полной потенциальной энергии балочного конечного элемента, узловые компоненты которого включают осредненные по высоте сечения углы трансверсального сдвига. Разработанная конечно-элементная модель протестирована на примерах решения классических задач. Ее актуальность показана расчетами на устойчивость монолитных изотропных балок с низкой трансверсальной сдвиговой жесткостью, трехслойных балок с податливым наполнителем, а также балок, изготовленных из однонаправленного композита (ортоаннизотропная модель). Полученные решения подтверждаются конечно-элементными расчетами, выполненными в среде пакета COSMOS/M в более строгой (двумерной) постановке.

Исследуемая конечно-элементная модель допускает описание граничных условий неклассического вида, когда на торцах балки регламентируется поведение углов трансверсального сдвига. Показано, что для некоторых типов граничных условий задача об устойчивости балки адекватным образом может быть решена только при использовании расчетных моделей, подобных описанной в данной работе.

Запишем выражение потенциальной энергии деформации в геометрически нелинейной постановке для задачи об изгибе в плоскости балки, податливой при трансверсальном сдвиге [1]:

$$U = \frac{1}{2} \int \int \int_{(V)} (S_x \Delta_x + S_z \Delta_z + S_{xz} \Delta_{xz} + S_{zx} \Delta_{zx}) dx dy dz, \quad (1)$$

где через S обозначены комплексы, связывающие напряжения

$$\begin{aligned} S_x &= \sigma_x + \tau_{xz} \theta_x, & S_{xz} &= \tau_{xz} - \sigma_x \omega_x, \\ S_z &= \sigma_z - \tau_{xz} \omega_x, & S_{zx} &= \tau_{xz} + \sigma_z \theta_x. \end{aligned}$$