

гласной договоренностью между создателями корпуса и его пользователями. Разрабатываемый в настоящее время Генеральный интернет-корпус русского языка (ГИКРЯ) задумывается как инструмент, позволяющий эксплицировать подобные договоренности и изучать русский язык в его дифференциальной полноте. Исследователи получают ресурс, позволяющий анализировать отдельные сегменты Интернета и создавать подкорпуса на основе метаразметки, извлекаемой автоматически. В настоящее время в ГИКРЯ размечены и доступны для поиска два сегмента русского Интернета: блог-платформа LiveJournal.com и «Журнальный зал». В дальнейшем количество сегментов планируется существенно расширить.

### **THE GENERAL INTERNET CORPUS OF RUSSIAN AND THE NOTION OF REPRESENTATIVENESS IN CORPUS LINGUISTICS**

**Piperski A.C.**

Russian State University for the Humanities, Institute of Linguistics, Moscow, Russia  
(Miuskaya Sq. 6-2, 125993, Moscow), e-mail: apiperski@gmail.com

The present article deals with the notion of representativeness in corpus linguistics. It turns out that there are no exact methods for assessing representativeness, and for this reason the representativeness of a corpus is nothing more than a tacit agreement between the creators of a corpus and its users. The General Internet Corpus of Russian (GICR) which is presently under development tries to make such an agreement explicit. It encourages its users to study register variation in the Russian language of the Internet. The linguistic community will be able to use a research tool to study different segments of the Web and to create subcorpora using automatically extracted metadata. As for June 2013, GICR contains two segments of the Russian Web, namely the blog platform LiveJournal.com and the "Magazine Reading Room" (<http://magazines.russ.ru/>). More segments will be added soon.

### **ЖАНРОВАЯ КЛАССИФИКАЦИЯ В ГЕНЕРАЛЬНОМ ИНТЕРНЕТ-КОРПУСЕ РУССКОГО ЯЗЫКА**

**Пиперски А.С.**

Институт лингвистики ФГБОУ ВПО «Российский государственный гуманитарный университет»,  
Москва, Россия (125993, Москва, Миусская пл., 6, корп. 2), e-mail: apiperski@gmail.com

Корпуса представляют собой важнейший инструмент современных лингвистических исследований. Для получения достоверных результатов исследователи, пользующиеся корпусами, должны обращать внимание на параметры метатекустовой разметки (информацию о социолингвистической, региональной, жанровой и т. п. принадлежности текста). В большинстве корпусов метатекустовые данные добавляются вручную, однако это невозможно при разработке больших корпусов, создаваемых на основе текстов из Интернета. Одним из таких корпусов является Генеральный интернет-корпус русского языка (ГИКРЯ), в котором применяются автоматические технологии метатекустовой разметки. В частности, предлагается новая схема жанровой разметки, при которой не выделяются априорные категории, а производится кластеризация на основе значений ряда переменных, выполняемая при помощи машинного обучения.

### **GENRE CLASSIFICATION IN THE GENERAL INTERNET CORPUS OF RUSSIAN**

**Piperski A.C.**

Russian State University for the Humanities, Institute of Linguistics, Moscow, Russia  
(Miuskaya Sq. 6-2, 125993, Moscow), e-mail: apiperski@gmail.com

Corpora are indispensable research tool in present-day linguistics. If a scholar wants to achieve reliable results in a corpus-based study, he should take into account metadata, i.e. sociolinguistic, regional and genre-related properties of the texts included into the corpus. In most corpora metadata are added manually, which is not possible when constructing large Web-based corpora. Since the General Internet Corpus of Russian (GICR) is one of such corpora, it has to use automated metadata tagging. The developers of GICR propose a novel approach to genre classification without postulating any a priori categories. Machine learning algorithms are used to cluster texts based on automatically extractable features.

### **МАКСИМАЛЬНО ПРАВДОПОДОБНАЯ ОЦЕНКА ДИСПЕРСИОННО-КОВАРИАЦИОННОЙ МАТРИЦЫ**

**Полянский И.С., Патронов Д.Ю.**

Академия ФСО России, Орел

В работе сформировано аналитическое выражение, определяющее максимально правдоподобную оценку дисперсионно-ковариационной матрицы наблюдения вектора случайных величин, распределенных по многомерному нормальному закону. Решение основано на получении выражения, определяющего точку экстремума сформированной на основе распределения Уишарта функции правдоподобия. Полученная оценка дисперси-

онно-ковариационной матрицы позволяет повысить эффективность оценки с точки зрения снижения разброса дисперсии ошибки. Что по существу приводит к увеличению точности оценки дисперсионно-ковариационной матрицы (на величину порядка нескольких тысяч раз) уже на выборках малого объема независимо от разрядности дисперсионно-ковариационной матрицы (размера вектора случайных величин). Причем вычислительные затраты сформированной оценки не хуже существующих. Представлены результаты экспериментальной проверки сформированного аналитического выражения максимально правдоподобной оценки дисперсионно-ковариационной матрицы.

#### **MAXIMUM REASONABLE ESTIMATE VARIANCE-COVARIANCE MATRICES**

**Polyanskiy I.S., Patronov D.Y.**

AcademyFSO of Russia, Orel

The work formed the analytic expression that defines the maximum likelihood estimate of variance-covariance matrix of the observation vector of random variables distributed according to multivariate normal distribution. The solution is based on obtaining an expression that identifies the extremum point formed on the basis of the distribution of Wishart likelihood function. The obtained estimate variance-covariance matrix to enhance the evaluation, in terms of reducing the spread of error variance. That substantially increases the accuracy of the estimate variance-covariance matrix (on the order of several thousand) already on the small sample size, regardless of the bit variance-covariance matrix (size of the vector of random variables). Moreover, the computational cost estimates formed better than the existing ones. The results of the experimental verification of the analytical expressions generated maximum likelihood estimation variance-covariance matrix.

#### **ВЛИЯНИЕ ЭФФЕКТОВ ОБТЕКАНИЯ НА ДИНАМИКУ ПРИРОДНОГО ПОЖАРА В УСЛОВИЯХ НЕОДНОРОДНОСТИ РЕЛЬЕФА**

**Постнов А.Д., Масленников Д.А., Катаева Л.Ю., Лоцилов С.А.**

ГОУ ВПО «Нижегородский государственный технический университет им. Р.Е. Алексеева», Нижний Новгород, Россия (603950, ГСП-41, Н.Новгород, ул. Минина, д.24), e-mail:nntu@nntu.nnov.ru

В статье рассматриваются вопросы, связанные с влиянием рельефа на динамику природного пожара за счёт эффектов обтекания. В основе проведённых расчётов лежит физическая модель лесных пожаров в двумерной постановке, учитывающая законы сохранения массы, импульса, энергии, а также химико-физические процессы. Для численного решения задачи используется метод Харлоу. В работе исследуется развитие пожара при различных конфигурациях рельефа: холм, равнина и овраг. Показаны изотермы и поля скоростей на различные моменты времени. Проводится анализ влияния рельефа на обтекание и как следствие на форму факела пламени и скорость пожара. Наглядно демонстрируется, что зависимость скорости пожара от рельефа не сводится к функции от угла наклона. Несмотря на то что на отдельных участках скорость пожара может существенно меняться, средняя скорость распространения приблизительно одинакова, что связано с малым углом наклона склонов холма и оврага ( $11^\circ$ ).

#### **INFLUENCE OF EFFECTS OF FLOW AROUND ON THE DYNAMICS OF NATURAL FIRE IN RELIEF IRREGULARITIES**

**Postnov A.D., Maslennikov D.A., Kataeva L.Y., Loshchilov S.A.**

1Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Nizhny Novgorod, Russia (603950, GSP-41, Nizhny Novgorod, ul. Minin St., 24)

The paper discusses issues related to the influence of topography on the dynamics of wildfire due to the effects of flow. The basis of the calculations is a physical model of forest fires in the two-dimensional formulation that takes into account the laws of conservation of mass, momentum, energy, and chemical and physical processes. For the numerical solution of the problem, a method of Harlow used. This paper investigates the propagation of a fire in different configurations of relief: hill, plain and ravine. Isotherm and the velocity field for various time points shown. The influence of relief on the flow and consequently on the shape of the flame and the speed of fire propagation analyzed. Clearly demonstrated the dependence relief on the fire rate propagation is not confined to the function of the angle of inclination. Despite the fact that in some areas the rate of fire can vary significantly, the average speed of propagation approximately the same, which is associated with a low angle slopes of the hill and ravine ( $11^\circ$ ).

#### **АВТОМАТИЧЕСКОЕ ВЫДЕЛЕНИЕ ИНФОРМАЦИИ ОБ АВТОРЕ И ИХ ТЕКСТАХ НА СТРАНИЦАХ ИНТЕРНЕТ-ФОРУМОВ**

**Пронин А.К., Копылов Н.Ю.**

ООО «Аби ИнфоПоиск», Москва, Россия (127273, Москва, ул. Отрадная, 2Б.6), e-mail: apronin@abby.com

В данной статье рассматривается метод автоматического выделения со страниц Интернет-форумов publicly доступной информации об авторе сообщений (пола, возраста, местоположения) и принадлежащих ему текстах. Для построения алгоритма использовалась концепция деревьев стилей, представляющих собой под-