

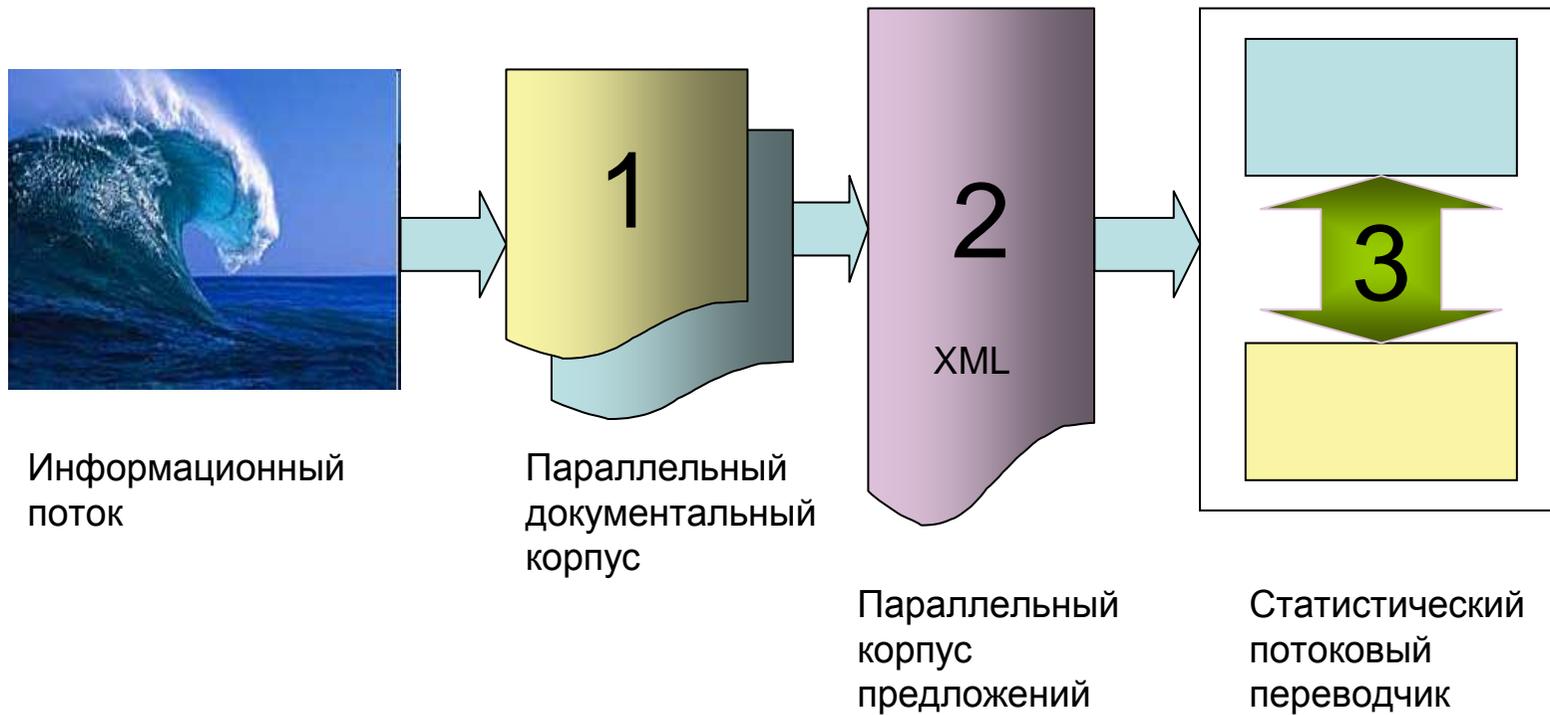


# ЭТАПЫ СОЗДАНИЯ СТАТИСТИЧЕСКОГО ПЕРЕВОДЧИКА ПОТОКОВ НОВОСТЕЙ

**Ландэ Дмитрий Владимирович,**  
д.т.н., профессор НТУУ «КТИ», зам. директора ELVisti  
**Жигало Владлен Викторович,**  
аспирант, инж.-программист ELVisti

СП6-2010

# Три задачи - три этапа





## Несколько слов о технологии

## КОНТЕНТ-МОНИТОРИНГА *InfoStream*

В Информационном центре "ЭЛВИСТИ" (Киев) создана система InfoStream, с помощью которой охватываются новости из более 5 тысяч отечественных и зарубежных веб-сайтов, осуществляется их обработка и обобщение.



## Основные характеристики



- более 5 000 российских и зарубежных источников;
- более 80 000 документов в сутки;
- архив с 1996 года содержит более 80 млн. документов;
- обновление данных осуществляется каждые 15 минут.





# 1. Параллельный корпус документов

Рассматривается метод, с помощью которого реализуется выявление информационных дубликатов, представленных на разных языках (русском и украинском).

В результате применения этого метода построен параллельный по информационному содержанию документальный корпус, который можно назвать «квазипараллельным», однако, он может также считаться параллельным в понимании многих авторов, так как оснащен некоторыми автоматически сформированными тегами и переводами выделенных лексем на 2 языка.



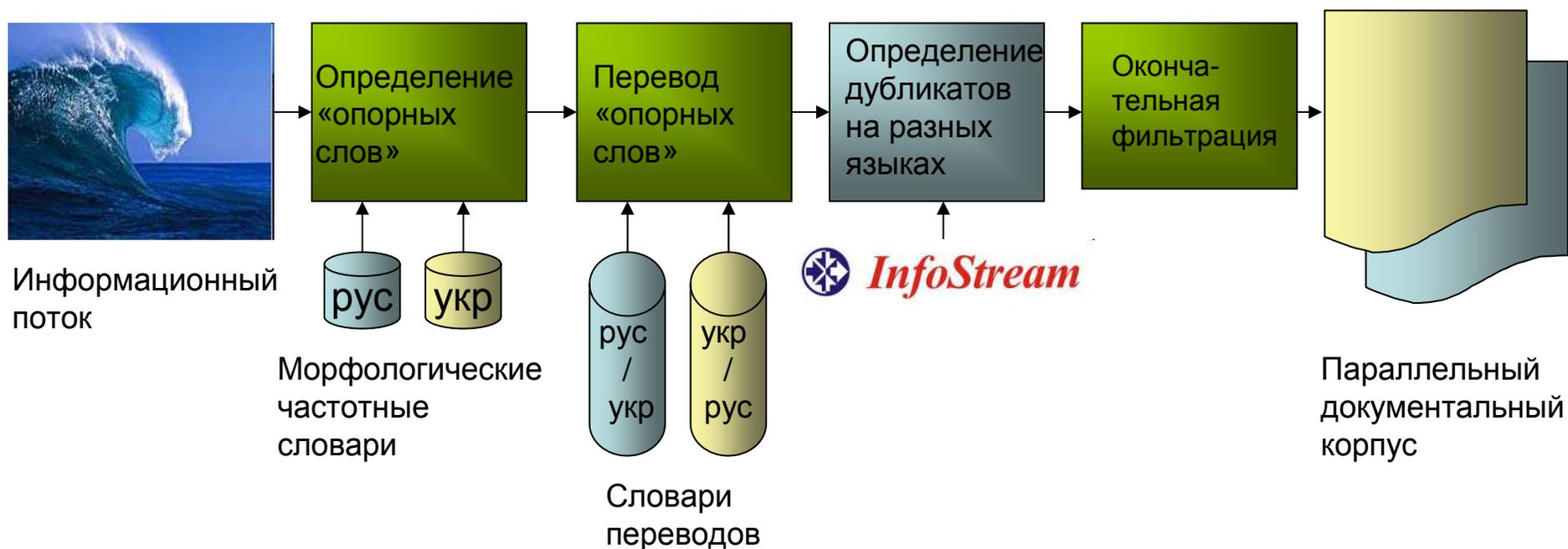
## Особенность подхода

Предлагается подход к созданию параллельных корпусов документов, основанный на алгоритме поиска дубликатов документов на разных языках. Подход дает возможность отыскать похожие документы на разных языках в большом массиве документов. В результате можно убедиться в том что в корпус попали параллельные документы из разных источников. Методы, основанные на анализе сайтов со страницами на разных языках, не позволяют определить дубликаты на разных источниках (сайтах), не указав специально параллельность этих источников.

# Процедура

1. Создание частотных морфологических словарей;
2. Выделение с их помощью опорных слов из документов;
3. Перевод опорных слов, с помощью словарей переводов;
4. Определение дублей документов на разных языках (сравнение 5-и переведенных опорных слов с 12 опорными словами др. документа);
5. Отсеивание с полученного множества документов «неполных дублей». Были использованы такие дополнительные критерии:
  - общее количество слов в переведенном варианте не должно отличаться больше чем на 10%;
  - количество слов начинающихся с большой буквы не должно отличаться больше чем на 3 слова;
  - количество чисел в документах не должно отличаться больше чем на два числа;
  - найденные числа в документах не должны отличаться более чем на 15 %.

# Алгоритм создания параллельного документального корпуса





## Что реализовано?

Предложенный подход позволил создать двуязычный украинско-русский параллельный корпус текстов из веб-публикаций на русском и украинском языках объемом около 1 000 000 пар документов. Оцененная экспертами точность предложенного алгоритма - 98%.

<http://ling.infostream.ua>



## Основные процедуры

Реализован алгоритм, который учитывает не только статистические свойства текстов, но и некоторые морфологические признаки.

В соответствии с этим алгоритмом построение параллельного корпуса происходит в несколько основных этапов:

- создание морфологических словарей (разово);
- создание частотных морфологических словарей (периодически);
- создание словарей переводов (постоянно);
- создание процедуры определения опорных слов в документах (постоянно);
- определение разноязычных дубликатов (постоянно).



## Морфологические словари

Для русского и украинского языков были использованы свободно доступные электронные словари: ispell с набором более 1 млн. словоформ и «Словники України», с набором более 4 млн. словоформ, а также словарь Зализняка, который насчитывает порядка 100 тыс. слов.

Эксперты дополнили морфологические словари неологизмами, названиями известных фирм, брендов и известными фамилиями, которых не было в исходных словарях.



## Частотные словари

Для обучения частотных морфологических словарей были взяты электронные публикации новостей, полученные из Интернет с помощью системы контент-мониторинга InfoStream.

«Обучение» словарей проводится в несколько этапов. Первый этап заключается в разделении документов на словоформы и сохранении полученных словоформ с информацией о номерах соответствующих документов.

На втором этапе, созданный файл словоформ сортируется, после чего подсчитывается количество вхождений каждой словоформы, и количество документов в которых она встретилась. Найденные частоты записываются в частотный словарь, на основании которого определяется вероятная нормальная форма каждого слова.

Для выявления омонимии, в выходной файл записываются все нормальные формы соответствующие словоформе, т. е. если одной словоформе соответствует сразу несколько нормальных форм, сохраняются подсчитанные частоты со всеми найденными нормальными формами. На третьем этапе происходит заключительный подсчет количества нормальных форм и сохранение результатов в частотный словарь.



## Контекстная неоднозначность

Для индексирования использовались украинско- и русскоязычные словарные массивы. Ввиду технической сложности представления полных лексикографических баз данных для двух языков, авторами использовался лишь относительно небольшой, но, по-видимому, самый существенный для данной задачи срез - множество имен существительных, дополненное некоторыми фамилиями, аббревиатурами, названиями компаний. Как показал опыт, такой подход полностью себя оправдал как для обеспечения качества индекса, так и для визуализации результатов работы.

Предложенный подход базируется на использовании частотного словаря на основе морфологического словаря (МС) с использованием тестового массива документов, а также построение алгоритма выявления опорных слов с использованием частотного МС и модификации общеизвестного подхода TF IDF.

## «Обучение» частотного словаря

Словоформа	Количество	Индекс нормальных форм
села	20	садиться → +20 село → +20
село	50	садиться → +50 село → +50
сели	10	садиться → +10
селом	30	село → +30
		село = 100 садиться = 80

## Опорные слова

Происходит считывание текстового документа из входного потока, после чего выполняется выделение словоформ и поиск нормальной формы для каждой из них. В случае контекстной неоднозначности, выбирается наиболее частотная (с наибольшим индексом) по словарю нормальная форма словоформы.

После вычисления соответствующих весовых коэффициентов с помощью формулы Окари VM25 происходит ранжирование нормализованных слов и выбирается двенадцать наиболее «весомых». Полученные двенадцать опорных слов переводятся на другой язык с помощью словарей переводов. Все опорные слова и слова-переводы приписываются к документу.

## Окарі ВМ25

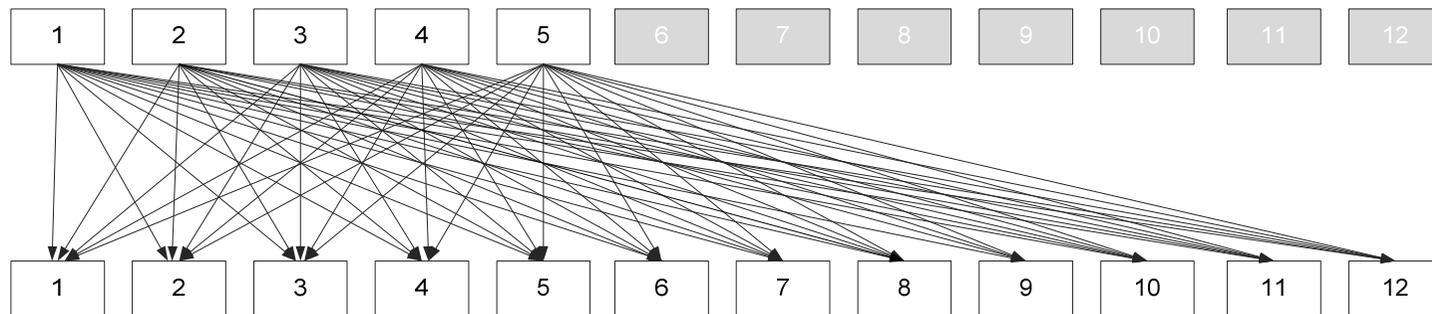
В предложенной процедуре индексирования для выделения наиболее значимых термов использовался статистический метод, базирующийся на применении общеизвестного подхода TF IDF, а точнее его модификации Окарі ВМ25, в которой каждому терму из документа приписывается вес по формуле:

$$W(t, D) = \frac{f(t, D)(k + 1)}{f(t, D) + k(1 - b + b|D|/L)} \cdot \log \frac{N - n(t) + 0.5}{n(t) + 0.5},$$

где  $f(t, D)$  - частота встречаемости терма  $t$  в документе  $D$ ,  $|D|$  - длина документа  $D$ ,  $L$  - **средняя длина документа в коллекции текстов**, общее количество которых -  $N$ ,  $n(t)$  - количество документов в коллекции, содержащих данный терм,  $k, b$  - параметры, выбираемые экспертами.

## Выявление дубликатов

В системе InfoStream используется механизм поиска дубликатов, в котором 6 опорных слов исследуемого документа, сравниваются с 12-ю опорными словами каждого из документов корпуса.



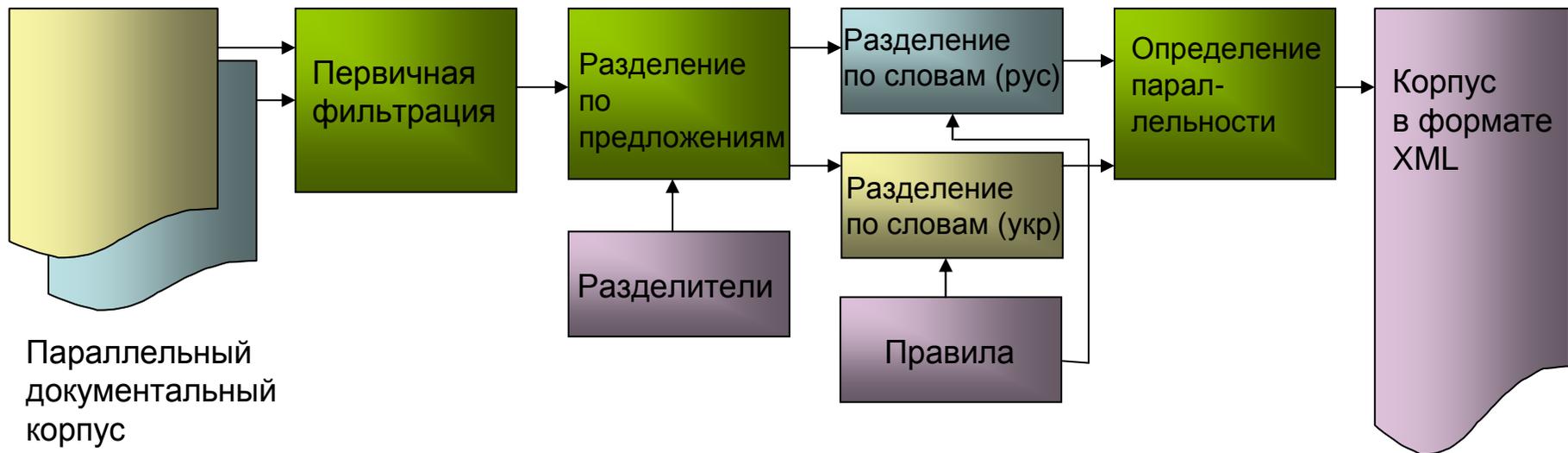
Процедура сравнения была дополнена рядом эвристических критериев, например:

- общее количество слов в переведенном варианте не должно отличаться от оригинала более чем на 10%;
- количество чисел в документах не должно отличаться больше чем на два.

## 2. Процедура создания корпуса параллельных предложений

1. Разделение параллельных документов на предложения:
  - 1.1. Определителем конца предложения были взяты символы (. ! ? ;)
  - 1.2. Если в тексте встречалось сокращение или инициалы с точкой то она не считается концом предложения.
2. Подсчет количества предложений в параллельных документах. Если данные документы по количеству предложений одинаковы, то они передавались в дальнейшую обработку.
3. Разделение предложения на слова:
  - 3.1. Словом считалось любое сочетание символов отделенное от других групп символов пробелом.
  - 3.2. Накладывались дополнительные ограничения на определение слова на каждом из языков. Например, слова на украинском, в начале которых, упоминались слова: який, яка, що, котрий и т.д. условно считались одним словом.
4. Подсчет количества слов в параллельных предложениях. В параллельный корпус предложений включались лишь те предложения, которые по количеству слов не отличись более чем на одно слово.

## Алгоритм создания корпуса параллельных предложений



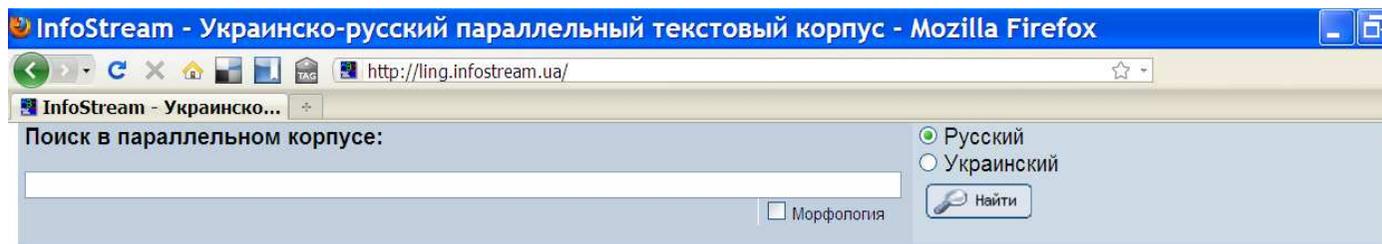
## Фрагмент параллельного корпуса

```

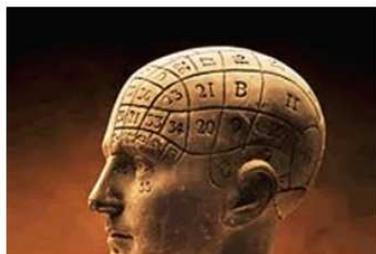
500-win.txt - Блокнот
Файл  Правка  Формат  Вид  Справка
<item>
    <rus>"Charge-Coupled Device" позволяет получать изображение,
преобразуя фотоны света в электроны (электрический ток).</rus>
    <ukr>"Charge-Coupled Device" дозволяє одержувати зображення,
перетворюючи фотони світла в електрони (електричний струм).</ukr>
</item>
<item>
    <rus>"Chevrolet" после столкновения раскрутило и он,
неуправляемый, задевает еще один "Nissan".</rus>
    <ukr>"Chevrolet" після зіткнення розкрутило і він, некерований,
зачіпає ще один "Nissan".</ukr>
</item>
<item>
    <rus>"Children-UA" это организация волонтеров из Европы и Канады,
которых объединяет неравнодушие к современной проблемы украинских
детей-сирот в Донецкой области.</rus>
    <ukr>"Children-UA" це організація волонтерів з Європи та Канади,
яких об'єднує небайдужість до сучасної проблеми українських дітей-сиріт в
Донецькій області.</ukr>
</item>
<item>
    <rus>"Chinese Democracy" станет первой полноценной пластинкой Guns
N'Roses с начала 1990-х годов.</rus>
    <ukr>"Chinese Democracy" стане першою повноцінною платівкою Guns
N'Roses з початку 1990-х років.</ukr>
</item>

```

# Онлайн-интерфейс - сайт <http://ling.infostream.ua>



## Украинско-русский параллельный текстовый корпус



В Информационном центре EIVisti создан выровненный на уровне предложений украинско-русский параллельный текстовый корпус из веб-публикаций. Объем корпуса - более 2,6 млн. пар уникальных предложений.

Метод построения корпуса базируется на использовании "опорных слов" в тестовых документах, а также средствах их

автоматического перевода. Опорные слова в рамках данного подхода выделяются с использованием русского и украинского морфологических словарей, а также словарей переводов имен существительных для русского и украинского языков. Кроме того, для вычисления весов терминов в документах используются некоторые дополнительные эмпирико-статистические правила. Для выравнивания параллельного корпуса на уровне предложений использовались преимущественно статистические методы.

Алгоритмы были реализованы в виде программного комплекса, который интегрирован с системой контент-мониторинга InfoStream, благодаря чему корпус постоянно пополняется.

Предполагается дальнейшее использование данного лингвистического ресурса для создания системы автоматического перевода новостных сообщений.

## Язык запросов

Запросы состоят из поисковых слов и операторов. В качестве поисковых слов могут использоваться слова естественного языка или их правые усечения. По умолчанию, при отключенной морфологии, каждое слово воспринимается как усечение (слова менее 3 символов ищутся как точное совпадение). Для поиска по полному слову, а не усечению, необходимо дописать к нему специальный символ "J". Система не чувствительна к регистрам букв.

В системе используется следующий набор операторов:

~ - оператор контекстного следования;

@ - оператор контекстной близости;

!, ^ - логическое И-НЕТ;

&, + - логическое И;

|, , - логическое ИЛИ;

Оператор контекстного следования (~) отбирает пары поисковых терминов, которые в тексте документа расположены друг за другом, причем учитывается порядок следования терминов.

Оператор контекстной близости (@) отбирает пары поисковых терминов, которые находятся рядом друг с другом, причем порядок следования не важен.

Различные уровни определяются с помощью круглых скобок.



# Описание ресурса - сайт <http://ling.infostream.ua>

Для скачивания доступен [заархивированный фрагмент параллельного корпуса](#) размером в **100 тысяч** пар уникальных предложений (в ZIP-архиве ~ 9 МБ).

Формат представления данных приближен к XML:

```
<item>
  <rus>предложение</rus>
  <ukr>речення</ukr>
</item>
... 99 998 раз ;)
<item>
  <rus>предложение</rus>
  <ukr>речення</ukr>
</item>
```

Информация представлена в кодировке CP1251 (Windows).

Использование этого фрагмента корпуса в научных и учебных целях - свободное.

Подробности - в [статье](#) Д.Ландэ и В.Жигало

Препринт: [arXiv:0807.0311](#), [PDF](#)

# Режим поиска - сайт <http://ling.infostream.ua>

Поиск в параллельном корпусе:

вертолет

Морфология

Русский  
 Украинский

↓ вертолет

Найдено документов - **3472**, страница 1 из 348

Статистика слов:

↓ **ВЕРТОЛЕТ** - 5651,

1. **Международная аэрокосмическая выставка "Фарнборо-2010" пройдет в Лондоне**

Около 1350 участников из 52 стран примут участие в международной аэрокосмической выставке "Фарнборо-2010", которая пройдет в пригороде Лондона 19-25 июля, сообщает BBC Russia со ссылкой на пресс-службу Федеральной службы по военно-техническому сотрудничеству (ФСВТС) России.

2. **"Си Бриз-2010": украинские десантники одели немцев в тельняшки**

В рамках учений "Си бриз 2010" в Украине немецкие десантники, за плечами которых не один десяток прыжков с парашютом, впервые осуществили прыжки с украинского вертолета Ми-8 под украинскими куполами.

3. **Пэрис Хилтон учится водить вертолет**

Пэрис Хилтон платит за один урок, на котором ее учат водить вертолет, 7 тысяч долларов.

4. **Под Берлином взрываются снаряды**

Борьба с огнем пока не очень успешна: пожарные не могут зайти в лес, чтобы приступить к тушению пожара, так как снаряды продолжают взрываться.

**Міжнародна аерокосмічна виставка "Фарнборо-2010" пройде в Лондоні**

Близько 1350 учасників з 52 країн візьмуть участь у міжнародній аерокосмічній виставці "Фарнборо-2010", яка пройде в передмісті Лондона 19- 25 липня, повідомляє BBC Russia з посиланням на прес-службу Федеральної служби з військово-технічного співробітництва (ФСВТС) Росії.

**"Сі Бриз-2010": українські десантники одягли німців у тільняшки**

У рамках вчень "Сі бриз 2010" в Україні німецькі десантники, за плечами яких не один десяток стрибків з парашютом, вперше здійснили стрибки з українського вертольота Мі-8 під українськими куполами.

**Періс Хілтон навчається водити гелікоптер**

Періс Хілтон платить за один урок, на якому її навчають водити гелікоптер, 7 тисяч доларів.

**Під Берліном вибухають снаряди**

Боротьба з вогнем поки що не дуже успішна: пожежники не можуть зайти в ліс, щоб приступити до гасіння пожежі, так як снаряди продовжують вибухати.

# Фрагмент целевого документа - сайт <http://ling.infostream.ua>

Документ по запросу: вертолет

## Международная аэрокосмическая выставка "Фарнборо-2010" пройдет в Лондоне

Около 1350 участников из 52 стран примут участие в международной аэрокосмической выставке "Фарнборо-2010", которая пройдет в пригороде Лондона 19-25 июля, сообщает BBC Russia со ссылкой на пресс-службу Федеральной службы по военно-техническому сотрудничеству (ФСВТС) России. "В статической экспозиции на открытых площадках будут демонстрироваться более 50 образцов авиационной техники из США, Канады, Великобритании, Италии, России, Украины.

Рядом с пассажирскими и спортивными самолетами будут демонстрироваться боевые самолеты F-18, F-15E, F-16, "Еврофайтер", военно-транспортный самолет C-27J, боевые **вертолеты** итальянской компании "Агуста Вестланд", говорится в сообщении.

[Закреть](#)

## Міжнародна аерокосмічна виставка "Фарнборо-2010" пройде в Лондоні

Близько 1350 учасників з 52 країн візьмуть участь у міжнародній аерокосмічній виставці "Фарнборо-2010", яка пройде в передмісті Лондона 19- 25 липня, повідомляє BBC Russia з посиланням на прес-службу Федеральної служби з військово-технічного співробітництва (ФСВТС) Росії. "У статичній експозиції на відкритих майданчиках демонструватимуться більше 50 зразків авіаційної техніки з США, Канади, Великобританії, Італії, Росії, України ...

Поряд з пасажирськими та спортивними літаками будуть демонструватися бойові літаки F-18, F-15E, F-16, "Єврофайтер", військово- транспортний літак C-27J, бойові вертольоти італійської компанії "Агуста Вестланд", говорить в повідомленні.

[Наверх](#)

# Примеры технологий n-gram

LDC 2006

Obtaining | Using | Providing | Creating Data

About LDC | Members | **Catalog** | Projects | Papers | LDC Online | Search / Help | Contact Us | UPenn | Home

LDC Catalog | [By Type and Source](#) | [By Year](#) | [Top Ten](#) | [Projects](#) | [Catalog Search](#)

## Web 1T 5-gram Version 1

Item Name: Web 1T 5-gram Version 1  
Authors: Thorsten Brants, Alex Franz  
LDC Catalog No.: LDC2006T13

Microsoft  
**Research**

Search Microsoft Research

Home | Our Research | **Collaboration** | Careers

About External Research | Tools and Services | Opportunities

Home > Collaboration > Our Focus > Computer Science > Microsoft Web N-gram Services

## Microsoft Web N-gram Services

Access *petabytes* of data via the [Public Beta Web N-gram Services](#).

We invite the whole community to use the Web N-gram services, made available via a cloud-based platform, to drive discovery and innovation in web-scale language processing, speech, and related areas by conducting research on real-world web-scale data, taking advantage of regular data updates that benefit from dynamic data.

The Web N-gram services provide you access to:

- Content types: Document Body, Document Title, Anchor Texts
- Model types: Smoothed models
- N-gram availability: unigram, bigram, trigram, N-gram with N=4, 5
- Training size (Body): All documents indexed by Bing in the en-us n-gram corpus
- Access: Hosted Services by Microsoft
- Updates: Periodical updates

Late last year, we introduced a private beta testing of the Web N-gram Services. We are now expanding access in the [Public Beta Web N-gram Services](#) to university professors, students, and researchers from around the world.

bing Web N-gram is brought to you by Microsoft Research in partnership with Microsoft Bing.

Google | Research Blog

## All Our N-gram are Belong to You

Thursday, August 03, 2006 at 8/03/2006 11:26:00 AM

Posted by Alex Franz and Thorsten Brants, Google Machine Translation Team

Here at Google Research we have been using word [n-gram models](#) for a variety of R&D projects, such as [statistical machine translation](#), speech recognition, [spelling correction](#), entity detection, information extraction, and others. While such models have usually been estimated from training corpora containing at most a few billion words, we have been harnessing the vast power of Google's datacenters and distributed processing [infrastructure](#) to process larger and larger training corpora. We found that there's no data like more data, and scaled up the size of our data by one order of magnitude, and then another, and then one more - resulting in a training corpus of *one trillion words* from public Web pages.

We believe that the entire research community can benefit from access to such massive amounts of data. It will advance the state of the art, it will focus research in the promising direction of large-scale, data-driven approaches, and it will allow all research groups, no matter how large or small their computing resources, to play together. That's why we decided to share this enormous dataset with everyone. We processed 1,024,908,267,229 words of running text and are publishing the counts for all 1,176,470,663 five-word sequences that appear at least 40 times. There are 13,588,391 unique words, after discarding words that appear less than 200 times.

Watch for an announcement at the Linguistics Data Consortium ([LDC](#)), who will be distributing it soon, and then order your set of 6 DVDs. And [let us hear from you](#) - we're excited to hear what you will do with the data, and we're always interested in feedback about this dataset, or other potential datasets that might be useful for the research community.

**Update (22 Sept. 2006):** The LDC now has the [data available](#) in their catalog. The counts are as follows:

File sizes: approx. 24 GB compressed (gzip'ed) text files

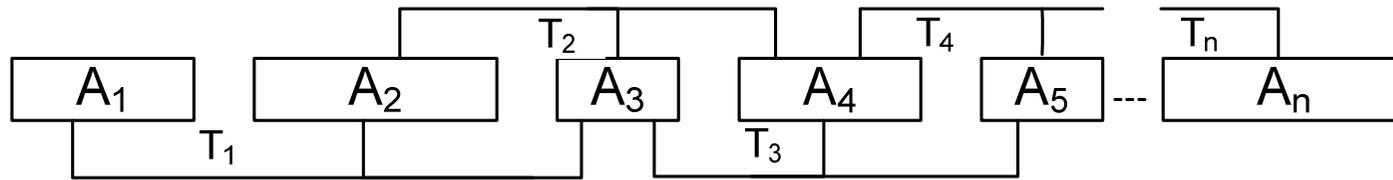
Number of tokens: 1,024,908,267,229  
Number of sentences: 95,119,668,584  
Number of unigrams: 13,588,391

- An Overview of Microsoft Web N-gram Corpus and Applications, NAACL-HLT 2010

Learn More

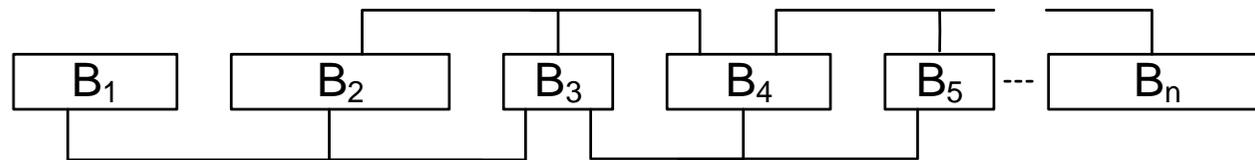
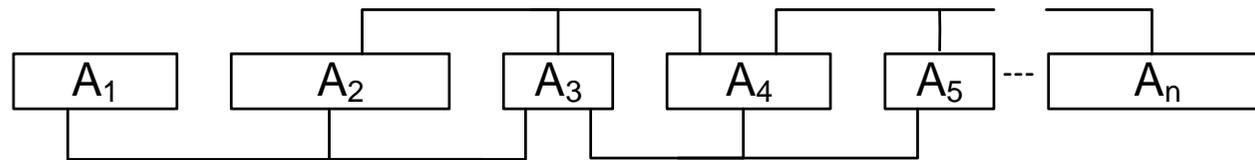
- Web N-gram FAQ
- Web N-gram Community Site

# Разбиение предложения на триграммы

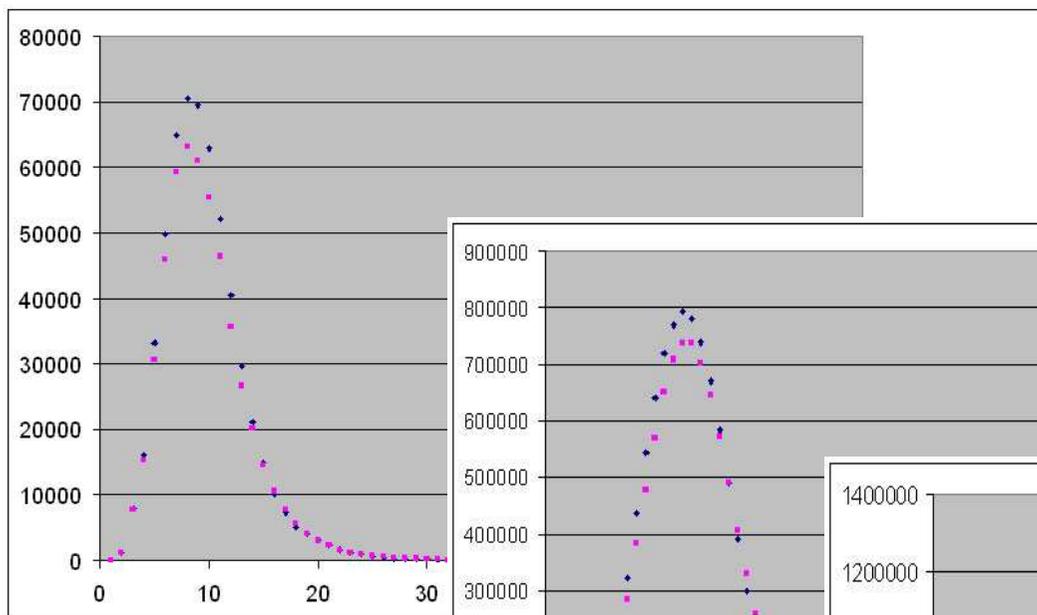


$A_1-A_n$  слова.  $T_1-T_n$  триграммы

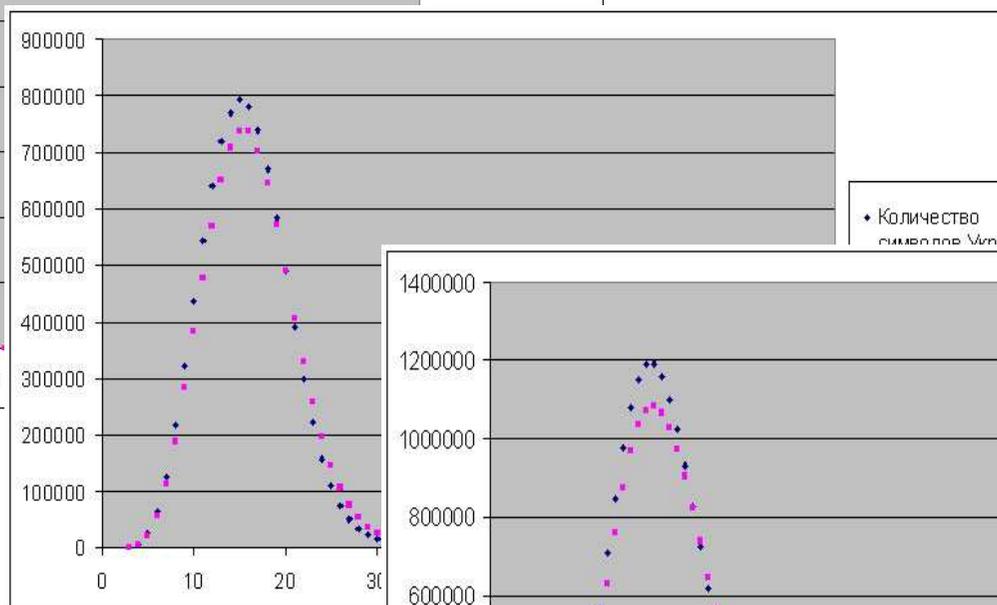
Предложения на русском и украинском языке.  
Схематическая разметка разбиения предложения на триграммы



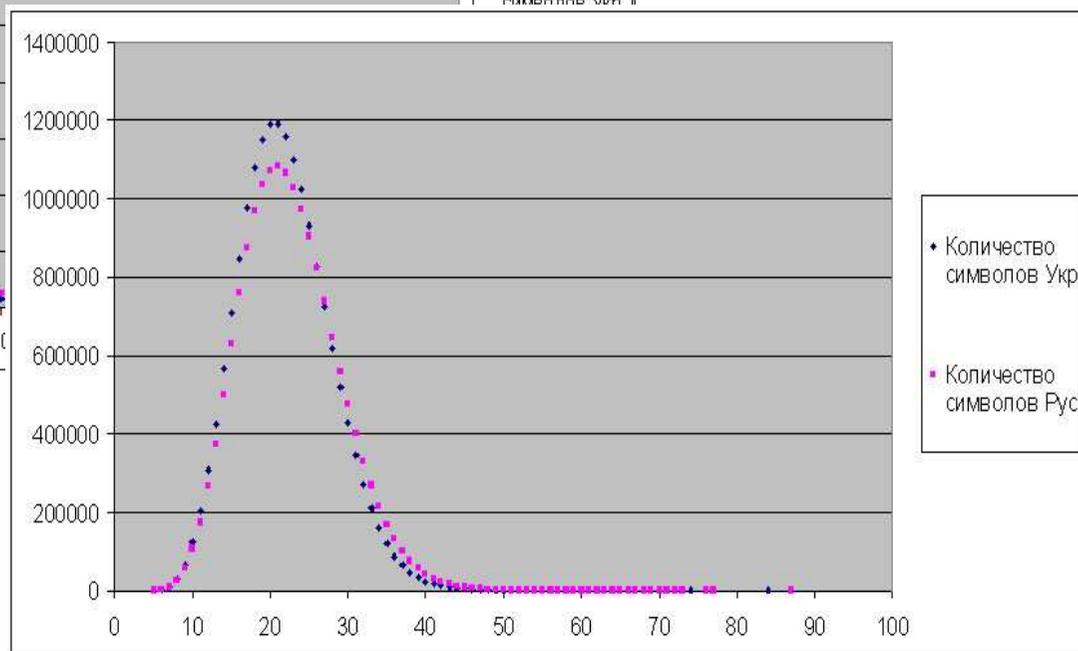
# Немного статистических исследований



2-граммы



3-граммы





## Статистика словарей

### Для русского языка

Триграммы:	16947925
Биграмммы:	8916423
Слова:	513753

### Для украинского языка

Триграммы:	17564031
Биграмммы:	9271080
Слова:	563876



## 3. Алгоритм работы переводчика

1. Разделение документов на предложения
2. Построение массивов триграмм, биграмм слов для документа
3. Поиск триграмм, биграмм и слов в словарях
4. Перевод документа с использованием построенных словарей переводов для документа
5. Форматирование документа

## Пример русско-украинского перевода

The image shows two windows from the AkelPad text editor. The top window, titled 'Рус.txt - AkelPad', contains the original Russian text. The bottom window, titled 'Укр.out - AkelPad', contains the translated Ukrainian text. The text describes the 200th anniversary of Alexander Pushkin in Lviv, including details about the 20th anniversary of the Russian Society named after Pushkin, the 210th anniversary of his birth, and various cultural events and exhibitions planned for June 6th and 7th.

**Рус.txt - AkelPad**

67 июня во Львове Пушкинские дни

67 июня во Львове будут отмечать Пушкинские дни. Все праздничные мероприятия будут посвящены 20-летию Русского общества им. А. Пушкина, основанного во Львове, и 210 годовщины со дня рождения великого русского поэта. Об этом сегодня, 2 июня, во время пресс-конференции сообщил председатель Русского общества им. А. Пушкина Олег Лютиков.

По его словам, 6 июня общественность проведет юбилейный Пушкинский поэтический праздник в селе Заболотивцы Жидачевского Львовской области. В этом селе находится памятник Пушкину, установленный еще в 1907 году, и именно вокруг этого памятника соберется молодежь, поэты и школьники, чтобы почтить память поэта.

7 июня во Дворце железнодорожников во Львове пройдет Юбилейный форум, посвященный этим двум датам. Участниками форума будут активисты общества, представители других организаций, представители органов власти. Приглашены гости из Червонограда, Стрия, Николаева, Ивано-Франковска и других городов. Почетным гостем станет Архиепископ Львовский и Галицкий Августин. Во время праздничного концерта будет выступать выдающаяся пианистка, лауреатка многих конкурсов Этелла Чуприк, которая исполнит произведения П. Чайковского и С. Рахманинова.

Также во время Пушкинских дней активисты Общества возложат цветы к памятникам Ивану Федорову, Тарасу Шевченко, Адаму Мицкевичу, на могилу Франко, к мемориальной доске Федорова в Свято-Онуфриевском монастыре.

Кроме того, запланирована выставка русских художников Львова.

13:71 | Ins | Unix | 21866 (украинская - КОИ8-U)

**Укр.out - AkelPad**

67 червня у Львові Пушкінські дні

67 червня у Львові будуть відзначати Пушкінські дні. Усі святкові заходи будуть присвячені 20-річчю Російського товариства ім. А. Пушкіна, заснованого у Львові, та 210 річниці з дня народження великого російської поета. Про це сьогодні, 2 червня, під час прес-конференції повідомив голова Російського товариства ім. А. Пушкіна Олег Лютиков.

За його словами, 6 червня товариство проведе ювілейне Пушкінське поетичне свято в селі Заболотівці Жидачівського Львівської області. У цьому селі знаходиться пам'ятник Пушкіну, встановлений ще в 1907 році, і саме навколо цього пам'ятника збереться молодь, поети та школярі, щоб вшанувати пам'ять поета.

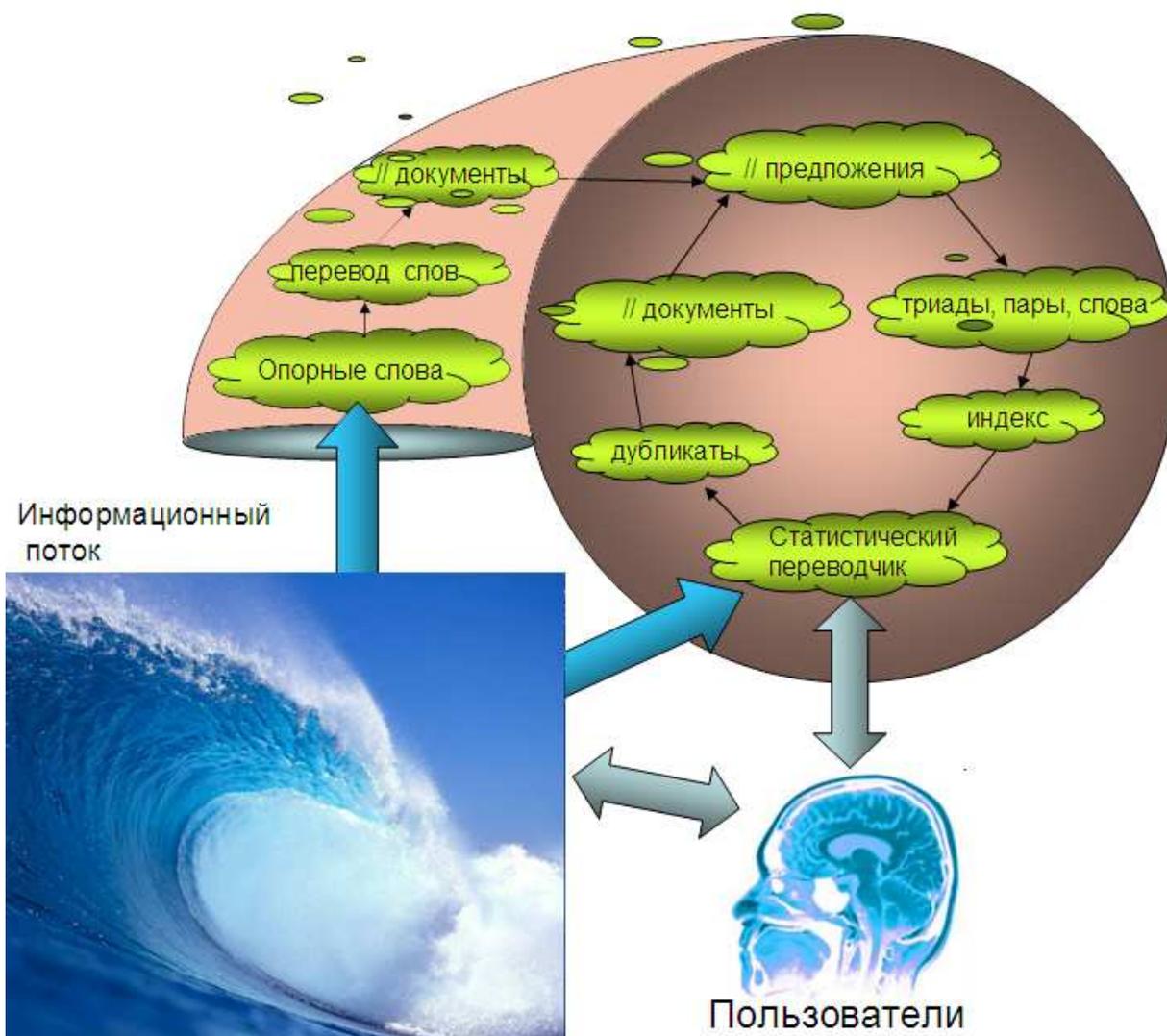
7 червня під Палаці залізничників у Львові пройде ювілейний форум, присвячений цих двох дат. Учасниками форуму будуть активісти товариства, представники інших організацій, представники органів влади. Запрошені гості з Червонограда, Стрия, Миколаєва, Івано-Франківська та інших міст. Почесним гостем стане архієпископ Львівський і Галицький Августин. Під час святкового концерту буде виступати видатна пианистка, Лауреатка багатьох конкурсів Етелла Чуприк, що виконає твори П. Чайковського і С.Рахманінова.

Також під час Пушкинских днів активісти Суспільства покладуть квіти до пам'ятників Івану Федорову, Тарасу Шевченку, Адаму Міцкевичу, на могилу Франка, до меморіальної дошки Федорова в Свято-Онуфриевском монастирі.

Крім того, запланована виставка російських художників Львова.

1:1 | Ins | Unix | 21866 (украинская - КОИ8-U)

# Место в технологии перевода ПОТОКОВ НОВОСТЕЙ





**СПАСИБО ЗА ВНИМАНИЕ!**

Ландэ Дмитрий Владимирович,  
[dwl@visti.net](mailto:dwl@visti.net)

<http://ling.infostream.ua>  
<http://dwl.visti.net>

