

DSPACE: ошибки, которых нужно избегать при создании и поддержке электронного архива.

Зональная Научная Библиотека
УРФУ

Александр Ефимов

<http://ideafix.name> ideafix@ideafix.name

2013

Типы ссылок DSPACE

- Ссылка на HANDLE – описание:

<http://aa.aa/handle/P/bb/>

- Ссылка на BITSTREAM – файл:

<http://aa.aa/handle/P/bb/N/X.Y>

Ссылка HANDLE

- <http://aa.aa/handle/P/bb/>
 - 1 - префикс протокола
 - 2 - доменное имя
 - 3 - указание на handle
 - 4 - значение префикса
 - 5 - конкретный handle

Ссылка BITSTREAM

• <http://aa.aa/handle/P/bb/N/X.Y>

1

- префикс протокола

2

- доменное имя

3

- указание на handle

4

- значение префикса

5

- номер handle

6

- номер bitstream

7

- имя документа

Источники ошибок

- `http://aa.aa/handle/P/bb/N/X.Y`
 - 1 - смена протокола (маловероятно)
 - 2 - Смена доменного имени
 - 3 - указание на handle
 - 4 - Смена префикса
 - 5 - Смена handle (крайне не желательно)
 - 6 - номер bitstream
 - 7 - имя документа (не желательно)

Источники ошибок

- I. Смена доменного имени
- II. Смена handle prefix
- III. Перенос фонда (смена сервера, восстановление бэкапа)
- IV. Изменение версии платформы.

«Уровни» актуализации ссылок

- I. Чтобы всё открывалось у нас
- II. Чтобы всё открывалось еще
и из поисковых систем
- III. Чтобы всё открывалось у
всех

Чтобы всё открывалось у нас

- Актуальные значения параметров **handle.canonical.prefix** и **handle.prefix**
- Верные редиректы на уровне WEB-сервера

Чтобы открывалось и из поисковых систем

- Актуальные значения в robots.txt
- Актуальный sitemap
- Ручное удаление неактуальных ссылок
- Ручная актуализация правил robots.txt

Чтобы открывалось отовсюду

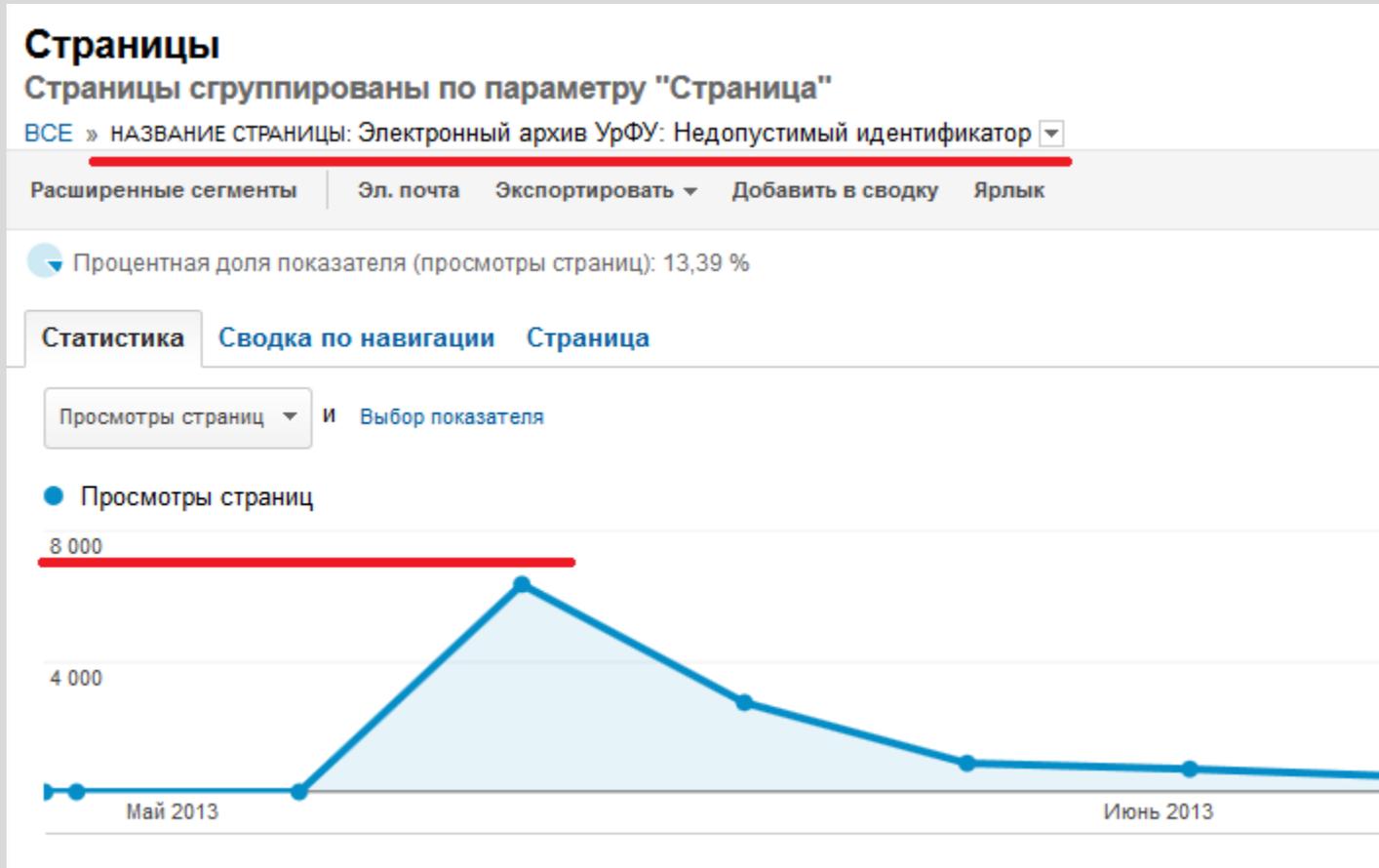
- Ручная актуализация правил редиректов WEB-сервера
- Поиск неактуальных входящих ссылок и их актуализация

Google Analytics, Яндекс Метрика

/bitstream/10995/1478/7/1324634_tasks.pdf	Электронный архив УрФУ: Недопустимый идентификатор
/bitstream/10995/1516/4/1331980_exam.pdf	Электронный архив УрФУ: Недопустимый идентификатор
/bitstream/10995/1601/5/1334887_schoolbook.pdf	Электронный архив УрФУ: Недопустимый идентификатор
/bitstream/10995/1724/8/1334956	Электронный архив УрФУ: Недопустимый идентификатор
/bitstream/10995/3963/1/conf_4-5.05.12.pdf	Электронный архив УрФУ: Недопустимый идентификатор

Это плохо, надо принимать меры

Google Analytics, Яндекс Метрика



Принятые меры работают!

Визуальный анализ проходимости

The image displays a visual analysis of website usability for the 'Электронный архив УрФУ' (Electronic Archive of Ural Federal University). It consists of two main parts: a Google Analytics dashboard and a heatmap overlaid on the website's navigation menu.

Google Analytics Dashboard

Показатель	Значение	В среднем по сайту
Просмотры страниц	2 719	9,07 % (29 982)
Уникальные просмотры страниц	2 014	9,72 % (29 725)
Средняя длительность просмотра страницы	00:00:58	В среднем по сайту: 00:00:42 (39,55 %)
Среднее время загрузки страницы (сек.)	4,06	В среднем по сайту: 1,47 (178,70 %)
Показатель отказов	68,72 %	В среднем по сайту: 71,20 % (-3,49 %)
Процент выходов	51,93 %	В среднем по сайту: 29,98 % (73,21 %)

Визуальный анализ (Heatmap)

The heatmap shows the interaction of users with the website's navigation elements. The most active areas are highlighted in red and orange, indicating high click-through rates.

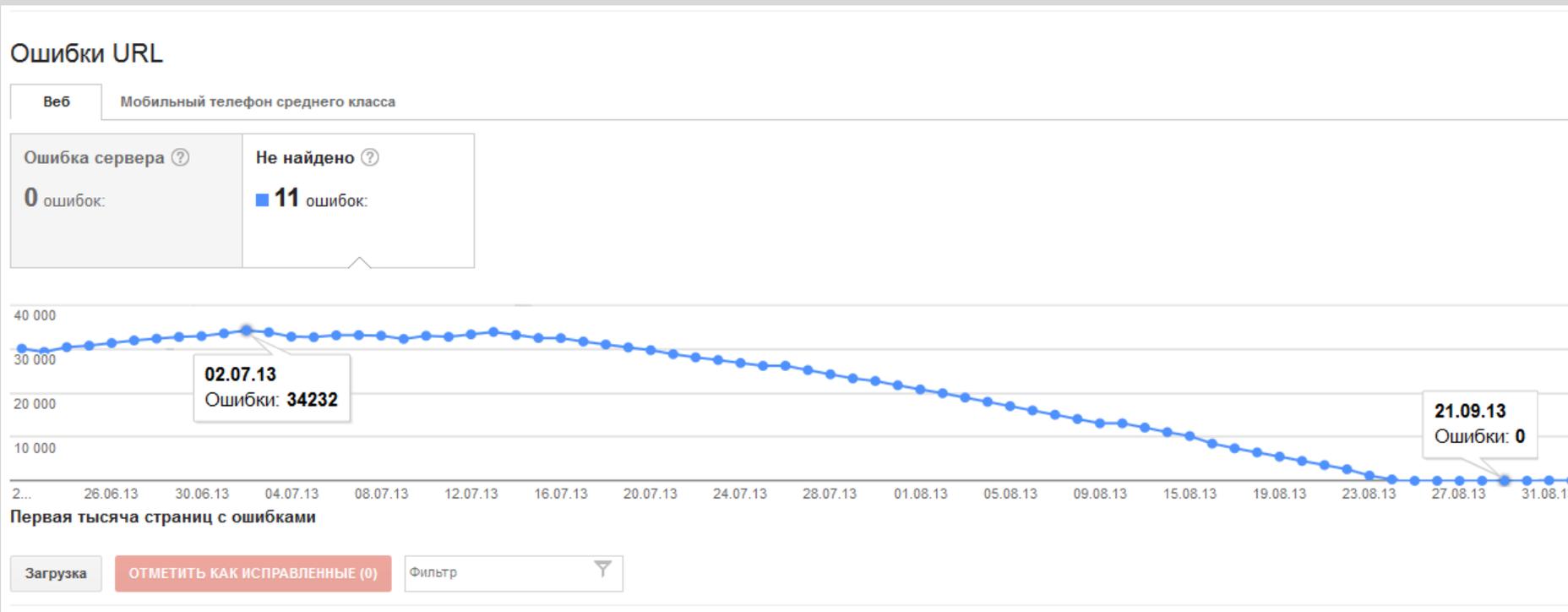
- Поиск в архиве:** The search bar and 'Найти' button are highly active.
- Разделы:** The top items in the list, such as 'Общегосударственные коллекции' and 'Высшая школа экономики и менеджмента', receive significant attention.
- Зарегистрированным:** The 'Обновления на e-mail' link is a key point of interaction.

Additional elements visible in the heatmap include the 'Главная страница' link, the 'Зональная научная библиотека УрФУ' link, and various institutional links in the footer.

WEBMASTER TOOLS

- GOOGLE WEBMASTER TOOLS (~35%)
- Яндекс ВЕБМАСТЕР (~50%)
- BING&YAHOO WEBMASTER TOOLS (<5%)

Google webmaster tools



Показывает общее количество ошибок и первую тысячу ошибочных ссылок.

Обновляется ежедневно!

Google webmaster tools

Не найдено

URL: <http://elar.usfeu.ru/bitstream/123456789/183/3/Kutiev>

Сведения об ошибке

В файле Sitemap

Ссылающиеся домены

<http://baza-dannuh.ru/r/dob-lap-hu-51.html>

Могут быть показаны не все URL.

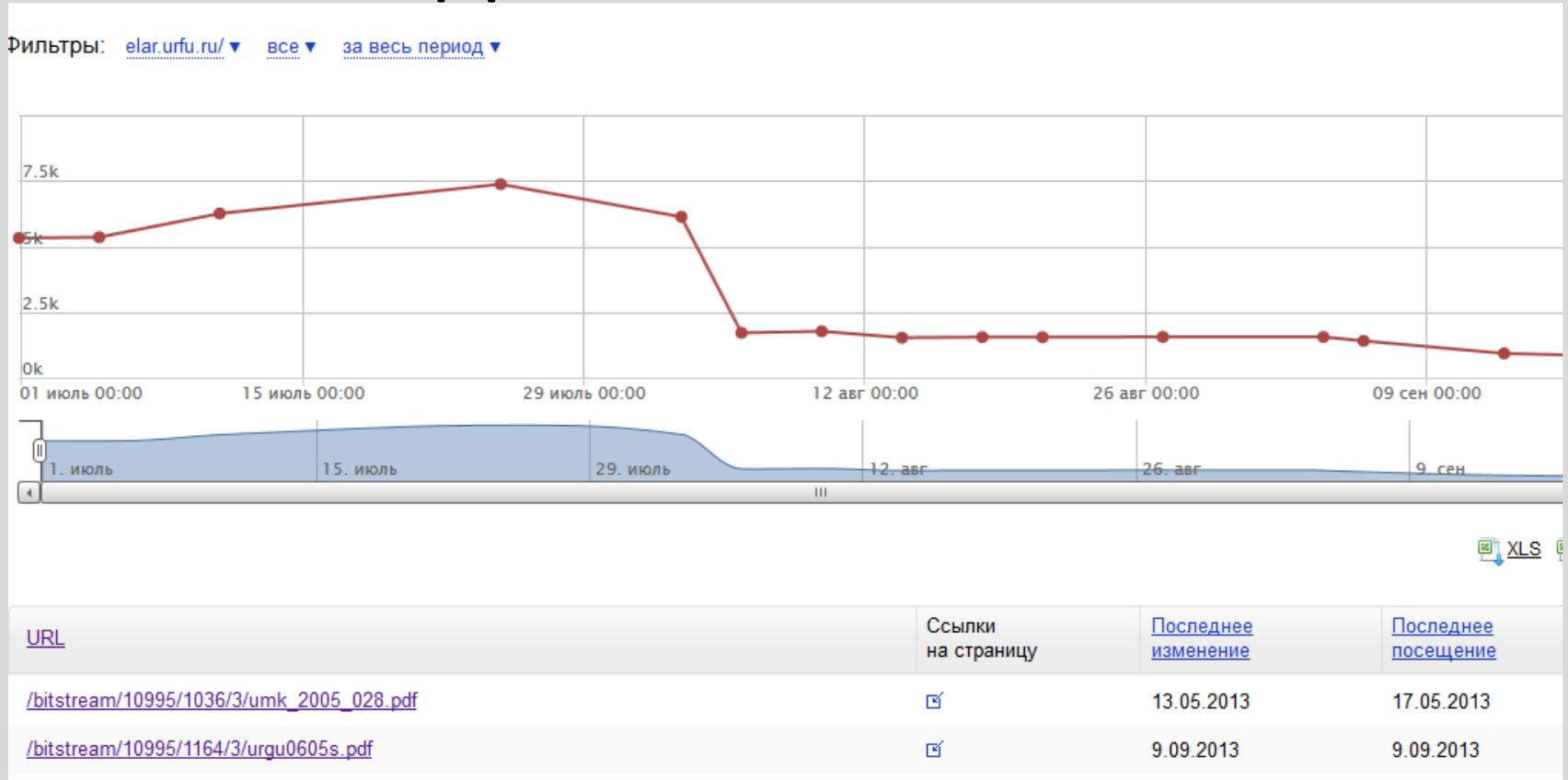
Отметить как исправленный

Отмена

The screenshot shows the website **beSMART** with a navigation menu (HOME, ABOUT, SERVICES, PROJECTS, CONTACT) and a search bar. The search results for the query "Dob hirdetés" are displayed, including a link to "Dob hirdetés - Több száz eladó használt és új ütőhangszer az Expressz.hu-n.", a blog post "Блог о путешествиях и культуре в Европе", and "DJs - миксы и треки". A sidebar on the right lists "ALSO LOOKING FOR" items like "Ночной шопинг в Дрим" and "Booking com Отель Avenida".

Позволяет узнать адрес размещения неактуальной ссылки и сделать выводы о целесообразности исправления.

Яндекс ВЕБМАСТЕР



Показывает общее количество ошибок и **все** ошибочные ссылки.

Обновляется примерно раз в две недели.

BING&YAHOO WEBMASTER TOOLS

АКТИВНОСТЬ НА САЙТЕ ⓘ

Сведения Bing о вашем сайте

	 ЩЕЛЧКИ ИЗ ПОИСКА	 ЕСТЬ В ПОИСКЕ	 СТРАНИЦЫ В ОБХОДЕ
% изменено	200% 	23 700% 	6 595% 
Данный период	2	237	2 812
Предыдущий период	0	0	42
Тренды			

Не очень удобный инструмент. Целесообразно использовать лишь для качественного представления содержимого архива в поисковых результатах BING и YAHOO

Инструменты на стороне сервера

- robots.txt
- sitemapS
- 301-302 редиректы

ROBOTS.TXT

ОБЯЗАТЕЛЬНЫЕ ДИРЕКТИВЫ

- **HOST** – имя хоста, особенно важно, когда есть зеркала и/или доступ по IP
- **SITEMAP** – путь к карте сайта
- **USERAGENT** – дабы случайно не отсеять поисковых ботов
- **ALLOW** – разрешенные для сканирования ссылки
- **DISALLOW** – запрещенные для сканирования ссылки

ROBOTS.TXT - пример

Sitemap: http://elar.urfu.ru/sitemaps/sitemap_index.xml.gz

Host: elar.urfu.ru

User-agent: *

Allow: /

Disallow: /browse-author?

Disallow: /browse-date?

Disallow: /browse-subject?

Disallow: /items-by-author?

Disallow: /items-by-subject?

Disallow: /browse-title?

Sitemaps

- Стандартный sitemap платформы – количество страниц для обхода равно количеству активных handle (количество заглавий + количество разделов + количество коллекций)
- Sitemap, сгенерированный внешними средствами – в зависимости от настроек может быть обойдена ВСЯ статика, количество страниц на ПОРЯДКИ больше чем в стандартном sitemap

Sitemaps - пример

URL-АДРЕС	ПОСЛЕДНЯЯ ОТПРАВКА	ОТПРАВЛЕННЫЕ URL-АДРЕСА	ТИП	ПОСЛЕДНИЙ ОБХОД	СОСТОЯНИЕ
<input type="checkbox"/> http://elar.usfeu.ru/sitemaps/sitemap_index.xml.gz	29.09.2013	1 205	Индекс карт сайтов	17.09.2013	Успешно
<input type="checkbox"/> http://elar.usfeu.ru/smbig/part03.gz	21.09.2013	33 007	Карта сайта	28.09.2013	Успешно
<input type="checkbox"/> http://elar.usfeu.ru/smbig/part02.gz	21.09.2013	50 000	Карта сайта	21.09.2013	Успешно
<input type="checkbox"/> http://elar.usfeu.ru/smbig/part01.gz	21.09.2013	50 000	Карта сайта	21.09.2013	Успешно

1 205 ссылок стандартного сайтмапа и **133 007** ссылок сгенерированного внешними средствами.

Постоянные и временные редиректы

- `UrlRewriteFilter` для Apache Tomcat
- `ModRewrite/.htaccess` для Apache httpd
- Действия на стороне DNS/NAT

UrlRewriteFilter *

Данный модуль является аналогом ModRewrite сервиса httpd, во многом повторяет его функционал, имеет xml based конфигурационные и ресурсные файлы, достаточно удобен в использовании и прост в установке и настройке.

* - Tomcat и редиректы - <http://ideafix.name/?p=991>

ModRewrite/.htaccess

Этот механизм возможно использовать только если Apache Tomcat работает бэкэндом за Apache httpd (при использовании vhost или mod_jk), настраивается на стороне httpd, прозрачен для Tomcat. В отличие от предыдущего метода, в некоторых случаях, для получения полной картины придется собирать и анализировать log'и в разных форматах из нескольких мест.

DNS/NAT

Служебные порты Tomcat, «старые» доменные имена и некоторые другие моменты можно настроить на стороне DNS сервера, либо фаерволла, находящегося между сервером с архивом и посетителями.

BONUS - Модификации JSPUI

Проблемы пожарной безопасности: пути их решения и совершенствование противопожарной защиты [Электронный ресурс] : материалы всероссийской научно-практической конференции с международным участием. – Екатеринбург: ФГАОУ ВПО «УрФУ имени первого Президента России Б.Н.Ельцина», 2012. – 163 с. – Режим доступа:

<http://elar.urfu.ru/handle/10995/4053>

BONUS - Модификации JSPUI

Черепанова, Е. С. Педагогика
толерантности [Электронный ресурс] :
прогр. повышения квалификации / Е.
С. Черепанова ; Урал. гос. ун-т им. А.
М. Горького. — Режим доступа:
http://elar.urfu.ru/bitstream/10995/1533/4/1332204_program.pdf. — (Дата
обращения: 26.09.2012).

BONUS - Модификации JSPUI

[jspui]/error/invalid-id.jsp

```
<li>Пояснительный текст (<script type="text/javascript">var a = window.location.href;var b = window.location.host;var result = a.match(/(\d+\d+)\V(.+)/);document.write("<a target='_blank' href='http://"+b+"/handle/" +result[1]+"'/>PURL</a>");</script>).</li>
```

Недопустимый идентификатор

Идентификатор 123456789/13/4/Ukazatel.pdf не соответствует правильному Поток битов архива электронных ресурсов. Это могло произойти по одной из следующих причин:

elar.usfeu.ru/handle/123456789/13

- Пожалуйста, используйте страницу описания ресурса для поиска файла [\(PURL\)](#).
- URL текущей страницы неверен. Если Вы попали сюда извне архива электронных ресурсов, то, возможно, адрес набран неправильно или поврежден.
- Вы ввели недопустимый ID в форму — пожалуйста, повторите попытку.

Если у Вас возникли проблемы или Вы считаете, что ID должен работать, то свяжитесь с администраторами сайта.

[Контакты администрации "Электронный архив УГЛТУ":](#)

[Перейти на главную страницу архива](#)

* - <http://ideafix.name/?p=1161>

BONUS – статические «заглушки»

```
<HTML>
```

```
<HEAD>
```

```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=UTF-8">
```

```
<META HTTP-EQUIV="Refresh" content="7; url=http://elar.usfeu.ru">
```

```
<TITLE>Внимание!</TITLE>
```

```
</HEAD><BODY>
```

```
<center>
```

```
<h1>Внимание!</h1>
```

```
<p>Ресурс более недоступен. Вы будете перенаправлены на главную страницу архива через 10 секунд.</p>
```

```
<p>Если Ваш браузер не поддерживает автоматические переходы, воспользуйтесь <a href="http://elar.usfeu.ru">ссылкой</a></p>
```

```
</center>
```

```
</BODY>
```

```
</HTML>
```

Внимание!

Ресурс более недоступен. Вы будете перенаправлены на главную страницу архива через 10 секунд.

Если Ваш браузер не поддерживает автоматические переходы, воспользуйтесь [ссылкой](http://elar.usfeu.ru)

Спасибо за внимание!

**Зональная Научная Библиотека
УРФУ**

Александр Ефимов

<http://ideafix.name> ideafix@ideafix.name

2013