

Приложение С.

Инструкция ассессора для дорожки Веб классификации

С.1 Постановка задачи

Целью работы является выявление множества Веб-страниц, которые содержат информацию на заданную тему.

Каждое задание состоит из:

- названия темы;
- поясняющего комментария;
- списка документов.

Технически задание распространяется в виде двух частей:

- Файла описания задания (с расширением .rtsk).
- Архива с набором документов для оценки (с расширением .bz2).

Результаты выполнения задания будут сохранены в файле с расширением .res.

С.2 Общие правила

Пожалуйста, придерживайтесь следующих правил при проведении оценки:

- Для работы с системой оценки ассессор должен зарегистрироваться и всегда использовать одно и то же имя пользователя. Не допускается использование одного и того же имени разными людьми.
- Задания следует выполнять последовательно. Параллельная работа над несколькими заданиями не желательна.
- Рекомендуются выполнять задания в том порядке, в каком система вам их показывает при выборе нового задания.
- Решение о степени соответствия страницы теме принимается вами самостоятельно в соответствии с вашим мнением. Однако вас могут попросить мотивировать ваш выбор.
- Пожалуйста, оценивайте сайты внимательно. Если в результате проверки будет выявлено низкое качество ваших оценок, то ваш труд может быть не зачтен.

С.3 Шкала оценки

Представьте, что вам необходимо составить подборку Веб страниц по заданной теме, выбирая их из некоторого множества кандидатов. Обратите внимание, что для ряда других тем есть отдельные разделы.

Для оценки сайтов используется следующая шкала:

- *Соответствующий (релевантный/витальный)*
На странице, безусловно, содержится много разносторонней информации по этой теме, и она в основном посвящена этой теме. Витальной может быть и страница с подборкой ссылок на ресурсы по данной теме, если набор ссылок подготовлен (структурирован, снабжен комментариями, а не выглядит, как набор случайных закладок) широко и подробно охватывает тему.
- *Скорее соответствующий (релевантный +)*
Документ содержит много полезной информации по заданной теме, но это не является главной темой или в значительной степени смешано с другой информацией. Например, документ затрагивает только специфическое подмножество вопросов.
- *Возможно соответствующий (релевантный -)*
Полезная информация по теме в тексте документа есть, но она явно частичная и не является основной темой страницы.
- *Не соответствующий (нерелевантный)*
Ничего полезного в тексте документа нет, или полезной информации крайне мало (например, заданная тема лишь косвенно упоминается).
- *Документ не может быть оценен*
Текст невозможно прочитать (представлен в некорректной кодировке или написан на непонятном языке), вызывает технические проблемы в браузере или не может быть оценен по каким-либо другим объективным причинам.

ВАЖНО: Обратите внимание, что Веб страница не обязательно должна относиться только к одной категории. Ситуация, когда представленная информация находится на стыке нескольких тем, вполне реальна, и в этом случае страница может быть отнесена к нескольким категориям.