

1. Entwurf

Es soll möglich sein, bei Erstellung einer Ressource zu deren Label eine transliterierte Variante zu erzeugen.

Da es zwei Transliterierungsrichtungen gibt (also entweder Deutsch/Englisch zu Kyrillisch oder umgekehrt), muss man die Voraussetzung annehmen, dass nur die Ressourcen betrachtet werden können, die mit einem Language-Label gekennzeichnet sind.

Von daher müssen wir zwei Fälle – das heißt sowohl deutsche als auch englische Namen bei der Transliterierung von der lateinischen in die kyrillische Schrift in Betracht ziehen.

Das Label bestimmt also die Transliterierungsrichtung.

Alles erfolgt über einen Transliterate-Button.

2. Dokumentation

Das ganze Verfahren beruht auf die Transliterierungsfunktion `transliterateRuString` (hier die Variante für kyrillisch-deutsch), die aus 4 Arrays mit entsprechenden Buchstaben in einer bestimmten Reihenfolge bestehen.

```
public function transliterateRuString($string){
    //...
    $del = array ('ju', 'Ju', 'JU', 'ja', 'Ja', 'JA',
'ch', 'Ch', 'CH', 'šč', 'Šč', 'ŠČ');
    $rul = array ('û',
'Û', 'Û', 'â', 'Â', 'Â', 'h', 'H', 'H', 's', 'S', 'S');
    $strans = str_replace($rul, $del, $string);
    $de2 = array
('\'', '´', 'A', 'B', 'C', 'D', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P',
'R', 'S', 'T', 'U', 'V', 'Y', 'Z',
'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'r', '
s', 't', 'u', 'v', 'y', 'z',
'Â', 'É', 'Û', 'â', 'è', 'û', 'č', 'č', 'š', 'š', 'š', 'š', 'ž', 'ž', '\\');
    $ru2 = array
('ъ', 'ь', 'А', 'Б', 'В', 'Д', 'Ф', 'Г', 'Х', 'И', 'Й', 'К', 'Л', 'М', 'Н', 'О',
'П', 'Р', 'С', 'Т', 'У', 'В', 'Ы', 'З', 'а', 'б', 'в', 'д', 'е', 'ф',
'г', 'х', 'и', 'й', 'к', 'л', 'м', 'н', 'о', 'п', 'р', 'с', 'т', 'у', 'в', 'ы', 'з', 'я',
'э', 'ю', 'я', 'э', 'ю', 'Ѳ', 'ѳ', 'Щ', 'щ', 'Ш', 'ш', 'Ж', 'ж', 'Ъ');
    $strans = str_replace($ru2, $de2, $string);
    return $strans;
}
```

Man könnte sich fragen, ob die 2 arrays - `$de1`, `$ru1` nicht überflüssig sind. Es dient aber einer besseren Bearbeitung der Daten – man markiert bei der Transliteration die Buchstaben, die aus 2 Lauten bestehen, z.B. dass Я nicht zu JA transliteriert wird, sondern zu Â. Dank dessen verliert man die Information nicht, falls man die Transliteration in die Rückrichtung ausführt, wird es in die richtige Я-Buchstabe umgewandelt.

Für die Ersetzung der Zeichen von einer Schrift in die andere benutzen wir die Funktion `str_replace`, der entsprechende Buchstabe wird in einem Array gefunden und wird durch das äquivalente Zeichen an der gleichen Stelle vom zweiten Array ersetzt.

```
$query->setProloguePart('SELECT ?label ')-
>setWherePart('WHERE{ <'. $resource.'> <http://www.w3.org/2000/01/rdf-
schema#label> ?label . '.
    'FILTER (langMatches(lang(?label), "de") ||
langMatches(lang(?label), "en") )'.
    '});
```

Die Transliterierung ist aber nur dann möglich, wenn die gewählte Ressource gelabelt ist.

```
$check->setProloguePart('SELECT ?label ')-
>setWherePart('WHERE{ <'. $resource.'> <http://www.w3.org/2000/01/rdf-
schema#label> ?label . '.
    'FILTER (langMatches(lang(?label), "dr"))'.
    '});

$this->model = $this->_owApp->selectedModel;
$queryResult = $this->model->sparqlQuery($query);
$checkResult = $this->model->sparqlQuery($check);
if ((is_array($checkResult) || isset
($checkResult[0]) || is_array($checkResult[0]))) {
    if (is_array($queryResult) || isset
($queryResult[0]) || is_array($queryResult[0])) {
        foreach($queryResult as $row){
            $this->_owApp->selectedModel-
>addStatement(
                $resource,
                'http://www.w3.org/2000/01/rdf-schema#label',
                array('value' => $this-
>transliterateDeString($row['label']), 'type' => 'literal', 'lang'
=> 'dr')
            );
        }
    }
}
```

Die zusätzliche Query `$check` in den verschachtelten if-Bedingungen verhindert die mehrfache Transliterierung von einem Label. Jede Transliterierung wird nämlich mit einem Label (entweder `dr` oder `rd`) versehen, so dass man weiß, dass der Prozess bereits stattgefunden hat und man mit einer transliterierten Variante eines Datensatzes zu tun hat.

3. Testsuite

- In den Arrays fehlen ein paar Buchstaben, wie z.B.: w, q, x, für die es keine Äquivalente im kyrillischen Alphabet gibt. Im Testfall kontrolliert man, ob, wenn ein Name einen von diesen Lauten enthält, dieser dann nach der Transliterierung auch noch vorhanden ist.
- Überprüfung, ob die erwähnten zweilautigen Buchstaben richtig umgewandelt werden
- Keine mehrfache Transliteration von dem gleichen Namen
- Ob jede transliterierte Variante wird mit einem entsprechenden Label gekennzeichnet wurde – ‘rd’ oder ‘dr’