Essential skills (3)

Tykushin Anatoly, a.tykushin@innopolis.ru, SNE

September 2016

Contents

1 URI parser

Task using **grep −E** utility parse html code of the page downloaded by **wget**. HTML code was retrieved from my own dokuwiki page.

To do this task the script was created. Script is listed below.

```
wget http://st5.os3.su:11180/doku.php && grep −E "(href|src|codebase|(blockquote|ins|q|del) cite|
profile|action|longdesc|usemap|classid|data|formaction|poster)=\"" doku.php > doku.php.parsed
```

Example of input file is shown in appendix A.

2 Dutch telephone number parser

Create a regexp to filter Dutch telephone land numbers correctly:

- with and without na7onal code, like 31205257414 — 0205257514

- with + or 00 like +31205257414 — 0031205257514

- With various spaces like +31 20 525 7514 — 00 31 525 75 14 — (020) 525 75 14

Regular expression, satisfying all requirements is shown below

```
^(((((\+|00|)\s?31\s?(20)?))+|(020|\(020\)))(\s?)(\d{7}|\d{3}\s((\d{2}\s?){2}|\d{4}))$
 1|----2----|---3---|--4----|------5------|--6--|---7--|---8---|---------9----------|10|
```

Where,

1. start of the line symbol

2. it can be either 00 and + symbols

3. it could be "00 31" "0031" "+31". Also it might be a space symbol in the end

4. after that could be "20" or nothing

5. either dutch number can start from 020 or (020)

6. after that may be any amount of space symbols, but I've limited their amount to 1, in order not to allow overflows

7. after that it might be 7 digits in a row

8. either 3 digits then a space

9. then 2 digits and space, 2 digits and space, or 4 digits

10. end of the line symbol

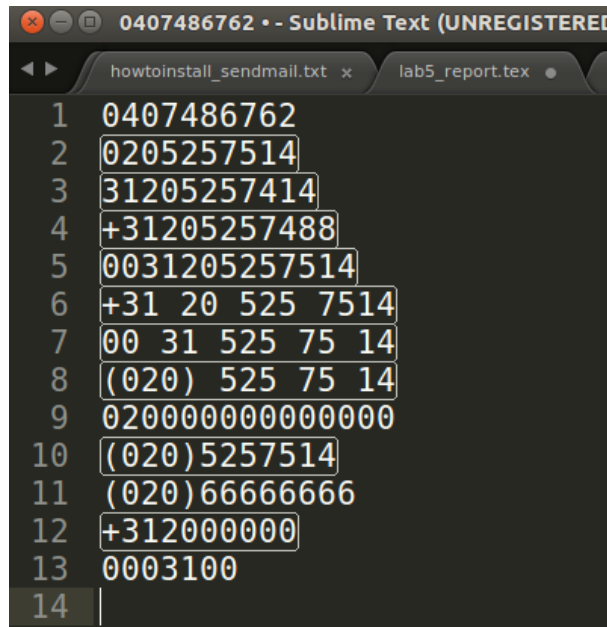On the figure 1 is shown how this regular expression works

Figure 1: Result of right choosing numbers

3 Matching VISA card

In official documents visa could have 13 or 16 digits

```
^(4([0-9]{3}))((\s|-)?)((([0-9]{4})\3){2})(?:[0-9]{3})?\d
|1|2|----3----|----4---|------5-----|4|-6-|------7-------|
```

Where,

1. start of string symbol

2. 4 means that it is visa paying system

3. 3 digits

4. here could be a delimiter: -, ' ', ''

5. 4 digits with/without the delimiters

6. 2 groups of them

7. 1 or 4 symbols (it depends on 13-digit or 16 digit VISA card number)

On the figure 2 is shown how this regular expression works

Figure 2: Result of right choosing numbers

4 Bash script comment eraser

```
grep -E '^([^#]|#!).*$' test_script.sh > test_script_after_regex.sh
      |1| 2  |3| 4 |
```

Where,

1. start of string symbol

2. string mustn't start with lone #

3. but it might start with #! (for example /bin/bash)

4. and other printable symbols

You can see example of how it works below:

```
file: test.sh                                   |    file: test_cut.sh

#!/bin/bash                                     |    #!/bin/bash
#fsadjf;s alfdjs                                |
s=`cat /usr/bin/unbound`    #this is            |    s=`cat /usr/bin/unbound`  # this is
# another comment                               |
dlkjas;                                         |    dlkjas;
fhsjdhfskf                                      |    fhsjdhfskf

#fdsafd                                         |
```

3

Appendices

A   URI's searching result in HTML

A.1   Input HTML markup.

It was got from dokuwiki main page

```
<!DOCTYPE html>
<html lang="en" dir="ltr" class="no-js">
<head>

    <meta charset="utf-8" />
    <title>start [ES_LS]</title>
    <script>(function(H){H.className=H.className.replace(/\bno-js\b/,'js')})(document.documentElemen
    t)</script>
    <meta name="generator" content="DokuWiki"/>
<meta name="robots" content="index,follow"/>
<meta name="keywords" content="start"/>
<link rel="search" type="application/opensearchdescription+xml" href="/lib/exe/opensearch.php"
title="ES_LS"/>
<link rel="start" href="/"/>
<link rel="contents" href="/doku.php?id=start&amp;do=index" title="Sitemap"/>
<link rel="alternate" type="application/rss+xml" title="Recent Changes" href="/feed.php"/>
<link rel="alternate" type="application/rss+xml" title="Current namespace"
href="/feed.php?mode=list&amp;ns="/>
<link rel="alternate" type="text/html" title="Plain HTML"
href="/doku.php?do=export_xhtml&amp;id=start"/>
<link rel="alternate" type="text/plain" title="Wiki Markup"
href="/doku.php?do=export_raw&amp;id=start"/>
<link rel="canonical" href="http://st5.os3.su:11180/"/>
<link rel="stylesheet" type="text/css"
href="/lib/exe/css.php?t=dokuwiki&amp;tseed=c81d1a85b02a91954dae8d56093319f9"/>
<script type="text/javascript">/*<![CDATA[*/var NS='';var JSINFO = {"id":"start","namespace":""};
/*!]]>*/</script>
<script type="text/javascript" charset="utf-8"
src="/lib/exe/js.php?t=dokuwiki&amp;tseed=c81d1a85b02a91954dae8d56093319f9"></script>
    <meta name="viewport" content="width=device-width,initial-scale=1" />
    <link rel="shortcut icon" href="/lib/tpl/dokuwiki/images/favicon.ico" />
<link rel="apple-touch-icon" href="/lib/tpl/dokuwiki/images/apple-touch-icon.png" />
    </head>

<body>
    <div id="dokuwiki__site"><div id="dokuwiki__top" class="site dokuwiki mode_show tpl_dokuwiki
    home  ">


<!-- ********** HEADER ********** -->
<div id="dokuwiki__header"><div class="pad group">


    <div class="headings group">
        <ul class="a11y skip">
            <li><a href="#dokuwiki__content">skip to content</a></li>
        </ul>

        <h1><a href="/doku.php?id=start"  accesskey="h" title="[H]"><img
        src="/lib/tpl/dokuwiki/images/logo.png" width="64" height="64" alt="" />
        <span>ES_LS</span></a></h1>
            </div>

    <div class="tools group">
```

```html
        <!-- USER TOOLS -->
                <div id="dokuwiki__usertools">
            <h3 class="a11y">User Tools</h3>
            <ul>
                <li><a href="/doku.php?id=start&amp;do=register"  class="action register"
                rel="nofollow" title="Register">Register</a></li><li><a
                href="/doku.php?id=start&amp;do=login&amp;sectok=14631a85f76200bcc878d38d86edbc3
                f"  class="action login" rel="nofollow" title="Log In">Log In</a></li>
                </ul>

        </div>


    <!-- SITE TOOLS -->
    <div id="dokuwiki__sitetools">
        <h3 class="a11y">Site Tools</h3>
        <form action="/doku.php?id=start" accept-charset="utf-8" class="search" id="dw__search"
        method="get" role="search"><div class="no"><input type="hidden" name="do" value="search"
        /><input type="text" placeholder="Search" id="qsearch__in" accesskey="f" name="id"
        class="edit" title="[F]" /><button type="submit" title="Search">Search</button><div
        id="qsearch__out" class="ajax_qsearch JSpopup"></div></div></form>              <div
        class="mobileTools">
            <form action="/doku.php" method="get" accept-charset="utf-8"><div class="no"><input
            type="hidden" name="id" value="start" /><select name="do" class="edit quickselect"
            title="Tools"><option value="">Tools</option><optgroup label="Page Tools"><option
            value="edit">Show pagesource</option><option value="revisions">Old
            revisions</option><option value="backlink">Backlinks</option></optgroup><optgroup
            label="Site Tools"><option value="recent">Recent Changes</option><option
            value="media">Media Manager</option><option
            value="index">Sitemap</option></optgroup><optgroup label="User Tools"><option
            value="login">Log In</option><option
            value="register">Register</option></optgroup></select><button
            type="submit">&gt;</button></div></form>               </div>
        <ul>
            <li><a href="/doku.php?id=start&amp;do=recent"  class="action recent" accesskey="r"
            rel="nofollow" title="Recent Changes [R]">Recent Changes</a></li><li><a
            href="/doku.php?id=start&amp;do=media&amp;ns="  class="action media" rel="nofollow"
            title="Media Manager">Media Manager</a></li><li><a
            href="/doku.php?id=start&amp;do=index"  class="action index" accesskey="x"
            title="Sitemap [X]">Sitemap</a></li>               </ul>
    </div>

</div>


<!-- BREADCRUMBS -->
        <div class="breadcrumbs">

                                <div class="trace"><span class="bchead">Trace:</span> <span
                                class="bcsep">●</span> <span class="curid"><bdi><a
                                href="/doku.php?id=start"  class="breadcrumbs"
                                title="start">start</a></bdi></span></div>
                </div>



    <hr class="a11y" />
</div></div><!-- /header -->

        <div class="wrapper group">


        <!-- ********** CONTENT ********** -->
        <div id="dokuwiki__content"><div class="pad group">
```

```html
        <div class="pageId"><span>start</span></div>

        <div class="page group">
                                                <!-- wikipage start -->
        <!-- TOC START -->
<div id="dw__toc">
<h3 class="toggle">Table of Contents</h3>
<div>

<ul class="toc">
<li class="level1"><div class="li"><a href="#essential_skills">Essential skills</a></div>
<ul class="toc">
<li class="level2"><div class="li"><a href="#assignments">Assignments</a></div>
<ul class="toc">
<li class="level3"><div class="li"><a href="#ways_to_go">Ways to go</a></div></li>
</ul></li>
</ul></li>
</ul>
</div>
</div>
<!-- TOC END -->

<p>
<a href="/lib/exe/fetch.php?media=logo-sne-big_white-inv_500x277.png" class="media"
title="logo-sne-big_white-inv_500x277.png"><img
src="/lib/exe/fetch.php?media=logo-sne-big_white-inv_500x277.png" class="mediacenter" alt="" /></a>
</p>

<h1 class="sectionedit1" id="essential_skills">Essential skills</h1>
<div class="level1">

</div>

<h2 class="sectionedit2" id="assignments">Assignments</h2>
<div class="level2">

</div>

<h3 class="sectionedit3" id="ways_to_go">Ways to go</h3>
<div class="level3">
<ul>
<li class="level1"><div class="li">     First topic is about <a href="/doku.php?id=fwc"
class="wikilink1" title="fwc">Smallest Federated Wiki</a> concepts</div>
</li>
<li class="level1"><div class="li">      Second topic is about creating <a
href="/doku.php?id=xml_xsd_dtd" class="wikilink1" title="xml_xsd_dtd">XML, XSD, DTD files</a></div>
</li>
<li class="level1"><div class="li"> Third topic is about creating XSLT file <a
href="/doku.php?id=xml_xsl_html_css" class="wikilink1" title="xml_xsl_html_css">XSL file</a></div>
</li>
</ul>

</div>

                <!-- wikipage stop -->
                                </div>

        <div class="docInfo"><bdi>start.txt</bdi> · Last modified: 2016/09/04 17:33 by
        <bdi>tyvision</bdi></div>
```

```
                    </div></div><!-- /content -->

            <hr class="a11y" />

            <!-- PAGE ACTIONS -->
            <div id="dokuwiki__pagetools">
                <h3 class="a11y">Page Tools</h3>
                <div class="tools">
                    <ul>
                        <li><a href="/doku.php?id=start&amp;do=edit"  class="action source"
                        accesskey="v" rel="nofollow" title="Show pagesource [V]"><span>Show
                        pagesource</span></a></li><li><a href="/doku.php?id=start&amp;do=revisions"
                        class="action revs" accesskey="o" rel="nofollow" title="Old revisions
                        [O]"><span>Old revisions</span></a></li><li><a
                        href="/doku.php?id=start&amp;do=backlink"  class="action backlink"
                        rel="nofollow" title="Backlinks"><span>Backlinks</span></a></li><li><a
                        href="#dokuwiki__top"  class="action top" accesskey="t" rel="nofollow"
                        title="Back to top [T]"><span>Back to top</span></a></li>
                    </ul>
                </div>
            </div>
        </div><!-- /wrapper -->


<!-- ********** FOOTER ********** -->
<div id="dokuwiki__footer"><div class="pad">
    <div class="license">Except where otherwise noted, content on this wiki is licensed under the
    following license: <bdi><a href="http://www.gnu.org/licenses/fdl-1.3.html" rel="license"
    class="urlextern">GNU Free Documentation License 1.3</a></bdi></div>
    <div class="buttons">
        <a href="http://www.gnu.org/licenses/fdl-1.3.html" rel="license"><img
        src="/lib/images/license/button/gnufdl.png" alt="GNU Free Documentation License 1.3" /></a>
        <a href="http://www.dokuwiki.org/donate" title="Donate" ><img
        src="/lib/tpl/dokuwiki/images/button-donate.gif" width="80" height="15" alt="Donate" /></a>
        <a href="http://php.net" title="Powered by PHP" ><img
            src="/lib/tpl/dokuwiki/images/button-php.gif" width="80" height="15" alt="Powered by
            PHP" /></a>
        <a href="http://validator.w3.org/check/referer" title="Valid HTML5" ><img
            src="/lib/tpl/dokuwiki/images/button-html5.png" width="80" height="15" alt="Valid HTML5"
            /></a>
        <a href="http://jigsaw.w3.org/css-validator/check/referer?profile=css3" title="Valid CSS"
        ><img
            src="/lib/tpl/dokuwiki/images/button-css.png" width="80" height="15" alt="Valid CSS"
            /></a>
        <a href="http://dokuwiki.org/" title="Driven by DokuWiki" ><img
            src="/lib/tpl/dokuwiki/images/button-dw.png" width="80" height="15" alt="Driven by
            DokuWiki" /></a>
    </div>
</div></div><!-- /footer -->

    </div></div><!-- /site -->

    <div class="no"><img src="/lib/exe/indexer.php?id=start&amp;1473765266" width="2" height="1"
    alt="" /></div>
    <div id="screen__mode" class="no"></div></body>
</html>
```

A.2 Here is the output file after grep -E command

```
<link rel="search" type="application/opensearchdescription+xml" href="/lib/exe/opensearch.php"
title="ES_LS"/>
<link rel="start" href="/"/>
<link rel="contents" href="/doku.php?id=start&amp;do=index" title="Sitemap"/>
<link rel="alternate" type="application/rss+xml" title="Recent Changes" href="/feed.php"/>
<link rel="alternate" type="application/rss+xml" title="Current namespace"
href="/feed.php?mode=list&amp;ns="/>
<link rel="alternate" type="text/html" title="Plain HTML"
href="/doku.php?do=export_xhtml&amp;id=start"/>
<link rel="alternate" type="text/plain" title="Wiki Markup"
href="/doku.php?do=export_raw&amp;id=start"/>
<link rel="canonical" href="http://st5.os3.su:11180/"/>
<link rel="stylesheet" type="text/css"
href="/lib/exe/css.php?t=dokuwiki&amp;tseed=c81d1a85b02a91954dae8d56093319f9"/>
<script type="text/javascript" charset="utf-8" src="/lib/exe/js.php?t=dokuwiki&amp;tseed=c81d1a85b02a9195
    <link rel="shortcut icon" href="/lib/tpl/dokuwiki/images/favicon.ico" />
<link rel="apple-touch-icon" href="/lib/tpl/dokuwiki/images/apple-touch-icon.png" />
        <li><a href="#dokuwiki__content">skip to content</a></li>
    <h1><a href="/doku.php?id=start"  accesskey="h" title="[H]"><img
    src="/lib/tpl/dokuwiki/images/logo.png" width="64" height="64" alt="" />
    <span>ES_LS</span></a></h1>
            <li><a href="/doku.php?id=start&amp;do=register"  class="action register"
            rel="nofollow" title="Register">Register</a></li><li><a
            href="/doku.php?id=start&amp;do=login&amp;sectok=14631a85f76200bcc878d38d86edbc3
            f"  class="action login" rel="nofollow" title="Log In">Log In</a></li>
            </ul>
        <form action="/doku.php?id=start" accept-charset="utf-8" class="search" id="dw__search"
        method="get" role="search"><div class="no"><input type="hidden" name="do" value="search"
        /><input type="text" placeholder="Search" id="qsearch__in" accesskey="f" name="id"
        class="edit" title="[F]" /><button type="submit" title="Search">Search</button><div
        id="qsearch__out" class="ajax_qsearch JSpopup"></div></div></form>            <div
        class="mobileTools">

            <form action="/doku.php" method="get" accept-charset="utf-8"><div class="no"><input
            type="hidden" name="id" value="start" /><select name="do" class="edit quickselect"
            title="Tools"><option value="">Tools</option><optgroup label="Page Tools"><option
            value="edit">Show pagesource</option><option value="revisions">Old
            revisions</option><option value="backlink">Backlinks</option></optgroup><optgroup
            label="Site Tools"><option value="recent">Recent Changes</option><option
            value="media">Media Manager</option><option
            value="index">Sitemap</option></optgroup><optgroup label="User Tools"><option
            value="login">Log In</option><option
            value="register">Register</option></optgroup></select><button
            type="submit">&gt;</button></div></form>            </div>
            <li><a href="/doku.php?id=start&amp;do=recent"  class="action recent" accesskey="r"
            rel="nofollow" title="Recent Changes [R]">Recent Changes</a></li><li><a
            href="/doku.php?id=start&amp;do=media&amp;ns="  class="action media" rel="nofollow"
            title="Media Manager">Media Manager</a></li><li><a
            href="/doku.php?id=start&amp;do=index"  class="action index" accesskey="x"
            title="Sitemap [X]">Sitemap</a></li>            </ul>
                        <div class="trace"><span class="bchead">Trace:</span> <span
                        class="bcsep">•</span> <span class="curid"><bdi><a
                        href="/doku.php?id=start"  class="breadcrumbs"
                        title="start">start</a></bdi></span></div>
<li class="level1"><div class="li"><a href="#essential_skills">Essential skills</a></div>
<li class="level2"><div class="li"><a href="#assignments">Assignments</a></div>
<li class="level3"><div class="li"><a href="#ways_to_go">Ways to go</a></div></li>
<a href="/lib/exe/fetch.php?media=logo-sne-big_white-inv_500x277.png" class="media"
title="logo-sne-big_white-inv_500x277.png"><img
src="/lib/exe/fetch.php?media=logo-sne-big_white-inv_500x277.png" class="mediacenter" alt="" /></a>
```

```
<li class="level1"><div class="li">        First topic is about <a href="/doku.php?id=fwc"
class="wikilink1" title="fwc">Smallest Federated Wiki</a> concepts</div>
<li class="level1"><div class="li">        Second topic is about creating <a
href="/doku.php?id=xml_xsd_dtd" class="wikilink1" title="xml_xsd_dtd">XML, XSD, DTD files</a></div>
<li class="level1"><div class="li"> Third topic is about creating XSLT file <a
href="/doku.php?id=xml_xsl_html_css" class="wikilink1" title="xml_xsl_html_css">XSL file</a></div>

                        <li><a href="/doku.php?id=start&amp;do=edit"  class="action source"
                        accesskey="v" rel="nofollow" title="Show pagesource [V]"><span>Show
                        pagesource</span></a></li><li><a href="/doku.php?id=start&amp;do=revisions"
                        class="action revs" accesskey="o" rel="nofollow" title="Old revisions
                        [O]"><span>Old revisions</span></a></li><li><a
                        href="/doku.php?id=start&amp;do=backlink"  class="action backlink"
                        rel="nofollow" title="Backlinks"><span>Backlinks</span></a></li><li><a
                        href="#dokuwiki__top"  class="action top" accesskey="t" rel="nofollow"
                        title="Back to top [T]"><span>Back to top</span></a></li>
                        </ul>
    <div class="license">Except where otherwise noted, content on this wiki is licensed under the
    following license: <bdi><a href="http://www.gnu.org/licenses/fdl-1.3.html" rel="license"
    class="urlextern">GNU Free Documentation License 1.3</a></bdi></div>
        <a href="http://www.gnu.org/licenses/fdl-1.3.html" rel="license"><img
        src="/lib/images/license/button/gnufdl.png" alt="GNU Free Documentation License 1.3" /></a>
        <a href="http://www.dokuwiki.org/donate" title="Donate" ><img
            src="/lib/tpl/dokuwiki/images/button-donate.gif" width="80" height="15" alt="Donate"
            /></a>
        <a href="http://php.net" title="Powered by PHP" ><img
            src="/lib/tpl/dokuwiki/images/button-php.gif" width="80" height="15" alt="Powered by
            PHP" /></a>
        <a href="http://validator.w3.org/check/referer" title="Valid HTML5" ><img
            src="/lib/tpl/dokuwiki/images/button-html5.png" width="80" height="15" alt="Valid HTML5"
            /></a>
        <a href="http://jigsaw.w3.org/css-validator/check/referer?profile=css3" title="Valid CSS"
        ><img
            src="/lib/tpl/dokuwiki/images/button-css.png" width="80" height="15" alt="Valid CSS"
            /></a>
        <a href="http://dokuwiki.org/" title="Driven by DokuWiki" ><img
            src="/lib/tpl/dokuwiki/images/button-dw.png" width="80" height="15" alt="Driven by
            DokuWiki" /></a>
    <div class="no"><img src="/lib/exe/indexer.php?id=start&amp;1473765266" width="2" height="1"
    alt="" /></div>
```