electronic document CONFERENCE

Deriving HTML from PDF

How reusable content in pdf could be

Roman Toda, Normex

History, Evolution of PDF

- Electronic representation of paper. Based on graphical model, accurate representation (screen and print)
  - Adopted word **rendering**
- Marked content
  - Introducing semantic into the content
- Tagged PDF
  - First time used different presentation model for assistive technology

Problems 2019

- PDF is essential part of the web – but is it really?
- HTML users hate pdf
  - How to use pdfs on mobiles
  - In browsers
- HTML developers hate pdf
  - Can’t control the user experience
  - Can’t access/navigate pdf content

Derivation

- Is it possible to deterministically interpret PDF content in the web based environment?
- Yes. **Tagged PDF** is the best way for capturing author’s intent
  - provisions in ISO 32000-2 are rich enough to be unambiguously interpreted in html = **Derived**
- **Derivation** our new word for **rendering**

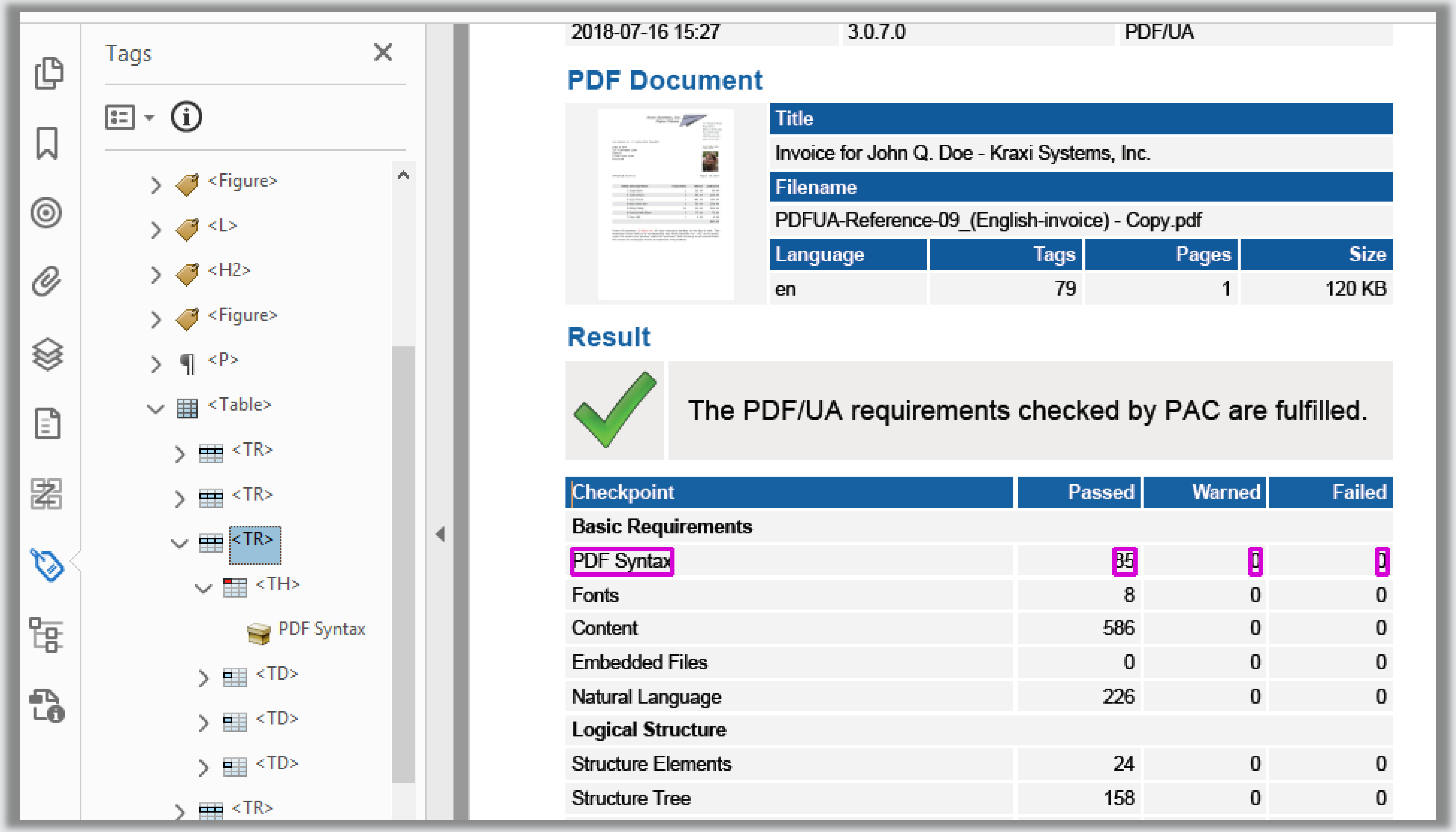
Join us

- Download and read the publication
- PDF Association
- Next-Generation PDF working group
- Test our MVP
  - https://github.com/Normex/PDF-Derivation
  - Share your files
  - Discuss techniques
  - Give feedback

What is Tagged PDF

- Structure tree
  - Marked content
  - Forms fields, links, annotations
- Attributes
- Classes
- Associated files
- Actions

What is Tagged PDF



2018-07-16 15:27 3.0.7.0 PDF/UA

PDF Document

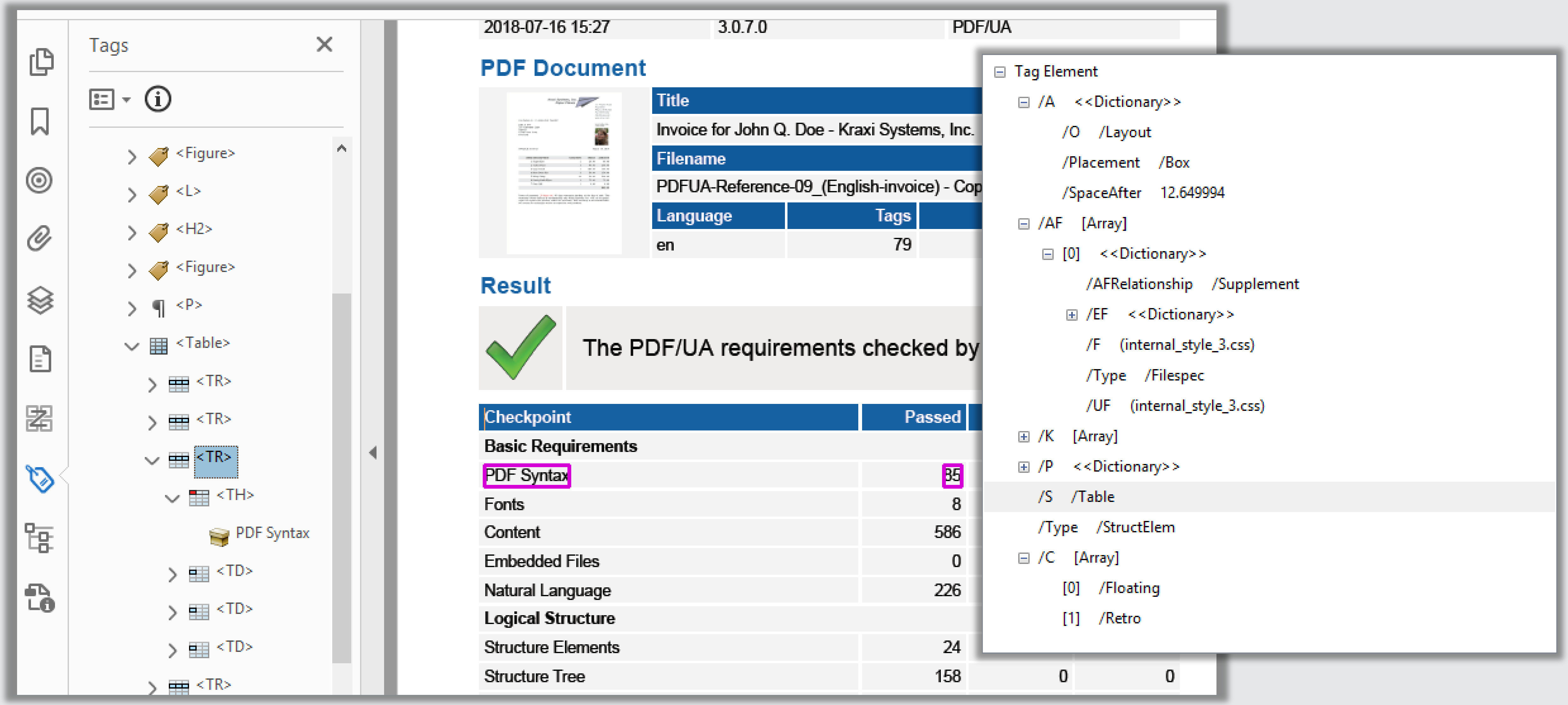
Title			
Invoice for John Q. Doe - Kraxi Systems, Inc.			
Filename			
PDFUA-Reference-09_(English-invoice) - Copy.pdf			
Language	**Tags**	**Pages**	**Size**
en	79	1	120 KB

Result

✓ The PDF/UA requirements checked by PAC are fulfilled.

Checkpoint	Passed	Warned	Failed
Basic Requirements			
PDF Syntax	85	0	0
Fonts	8	0	0
Content	586	0	0
Embedded Files	0	0	0
Natural Language	226	0	0
Logical Structure			
Structure Elements	24	0	0
Structure Tree	158	0	0

What is Tagged PDF



```
> <Figure>
> <L>
> <H2>
> <Figure>
> ¶ <P>
v <Table>
  > <TR>
  > <TR>
  v <TR>
    v <TH>
        PDF Syntax
    > <TD>
    > <TD>
    > <TD>
  > <TR>
```

2018-07-16 15:27	3.0.7.0	PDF/UA

PDF Document

Title	
Invoice for John Q. Doe - Kraxi Systems, Inc.	
Filename	
PDFUA-Reference-09_(English-invoice) - Cop	
Language	Tags
en	79

Result

✓ The PDF/UA requirements checked by

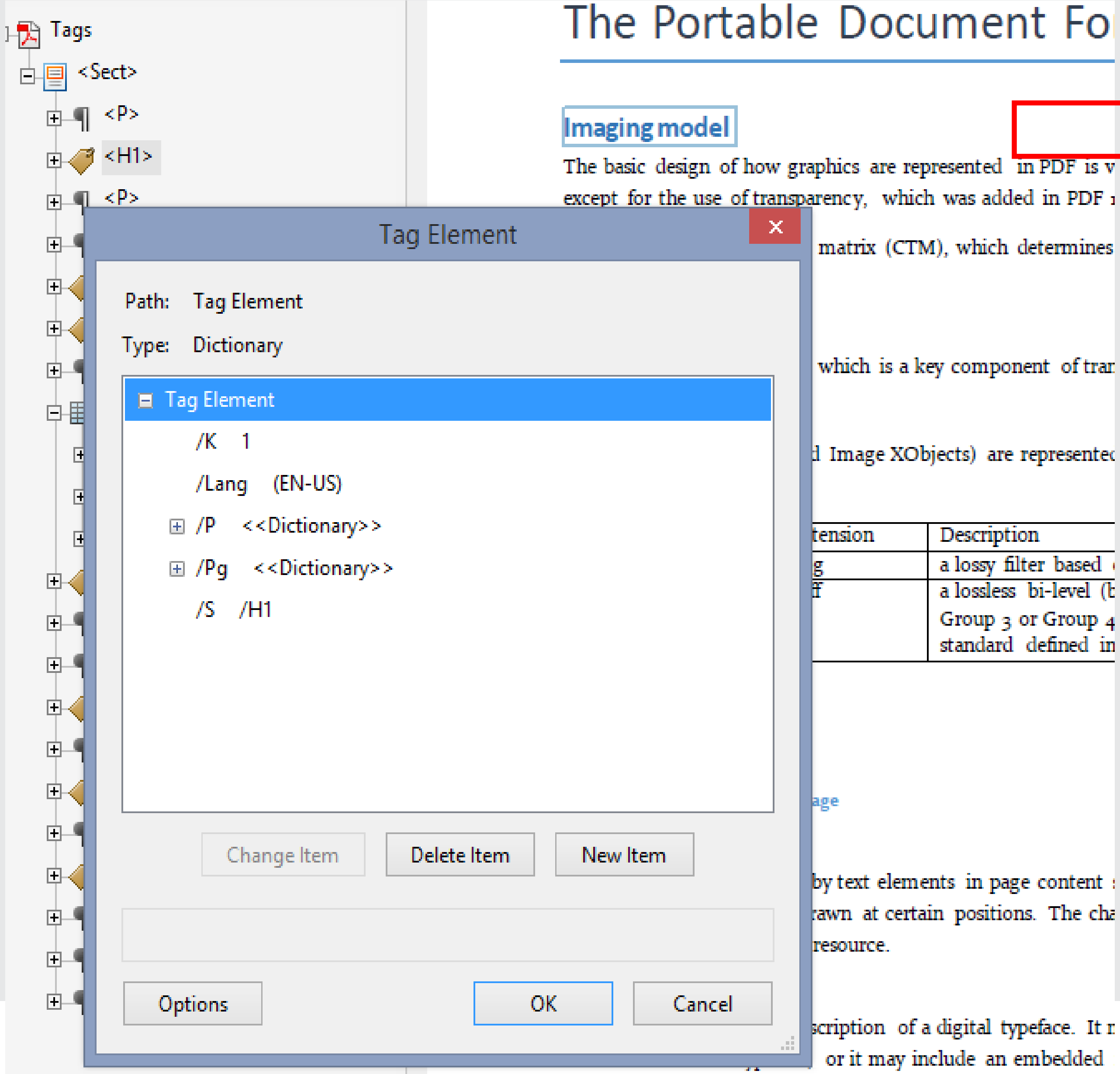
Checkpoint	Passed		
Basic Requirements			
PDF Syntax	85		
Fonts	8		
Content	586		
Embedded Files	0		
Natural Language	226		
Logical Structure			
Structure Elements	24		
Structure Tree	158	0	0

```
⊟ Tag Element
   ⊟ /A     <<Dictionary>>
        /O      /Layout
        /Placement    /Box
        /SpaceAfter    12.649994
   ⊟ /AF    [Array]
      ⊟ [0]    <<Dictionary>>
           /AFRelationship    /Supplement
         ⊞ /EF    <<Dictionary>>
           /F    (internal_style_3.css)
           /Type    /Filespec
           /UF    (internal_style_3.css)
   ⊞ /K    [Array]
   ⊞ /P    <<Dictionary>>
     /S    /Table
     /Type    /StructElem
   ⊟ /C    [Array]
        [0]    /Floating
        [1]    /Retro
```

Standard structure element types

- Chapter 14.8.4 in ISO 32000-2
  - Annot, Artifact, Aside, Caption, Document, DocumentFragment, Div, Em, FEnote, Figure, Form, Formula, H1..Hn, L, Lbl, LBody, LI, Link, P, RB, RP, RT, Ruby, Span, Strong, Sub, Table, TBody, TD, TFoot, TH THead, Title, TR, Warichu, WT, WP
- Some sound like html tags: Div, Span, P, H1, LI
- Some are harder to imagine: Annot
- RoleMap may be involved

Text



The Portable Document Fo

Imaging model

The basic design of how graphics are represented in PDF is v
except for the use of transparency, which was added in PDF
matrix (CTM), which determines
which is a key component of tran
Image XObjects) are represented

tension	Description
g	a lossy filter based
f	a lossless bi-level (b Group 3 or Group 4 standard defined in

by text elements in page content
rawn at certain positions. The cha
resource.

scription of a digital typeface. It r
or it may include an embedded

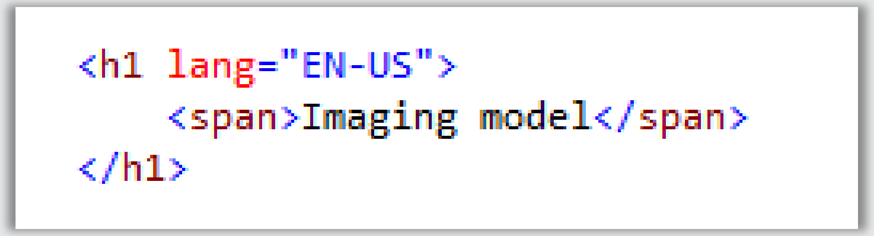
Tag Element

Path: Tag Element

Type: Dictionary

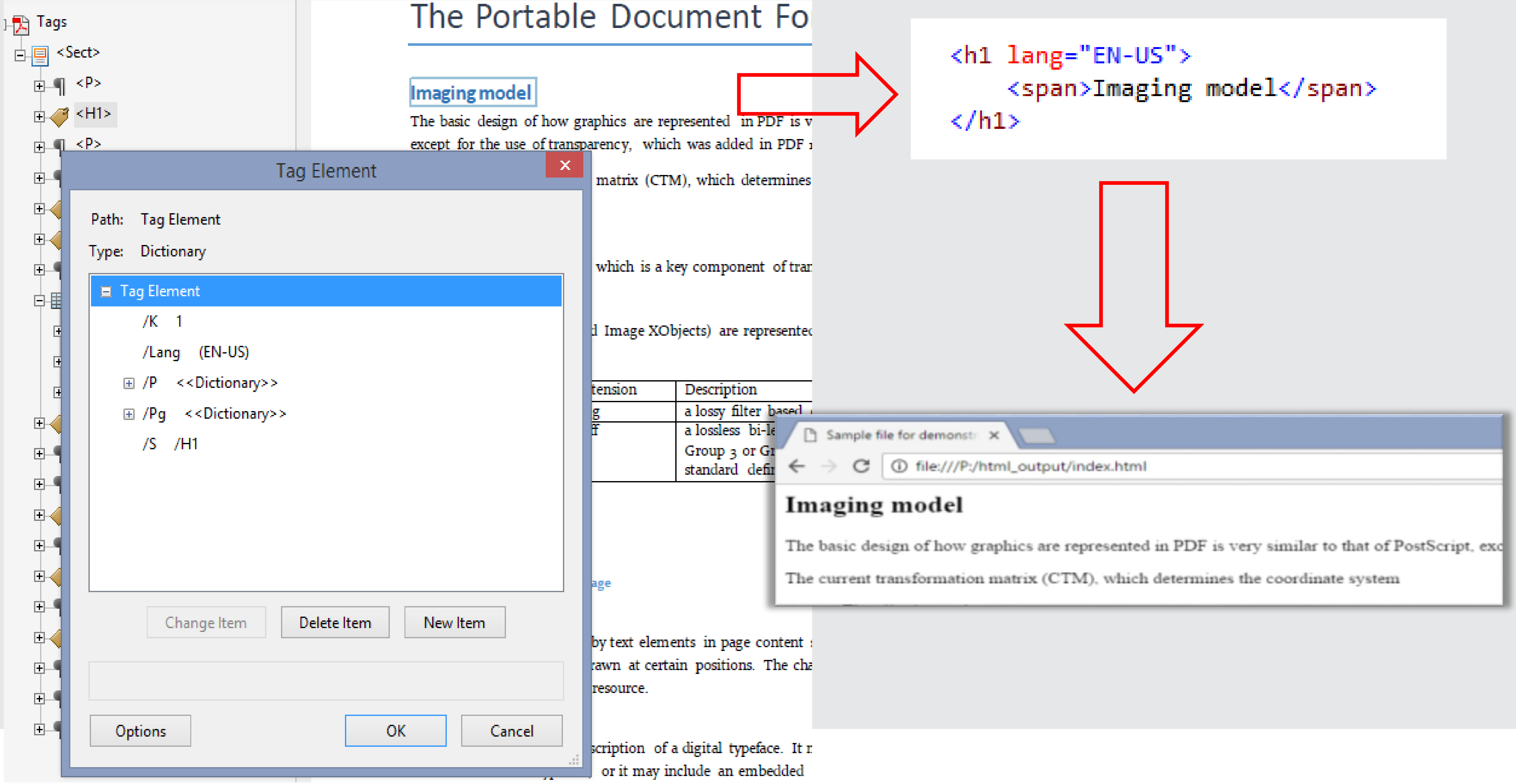
```
⊟ Tag Element
    /K     1
    /Lang    (EN-US)
  ⊞ /P     <<Dictionary>>
  ⊞ /Pg    <<Dictionary>>
    /S    /H1
```

Change Item
Delete Item
New Item
Options
OK
Cancel



```
<h1 lang="EN-US">
    <span>Imaging model</span>
</h1>
```

Simple structure - Text



Tag Element

Path: Tag Element

Type: Dictionary

```
⊟ Tag Element
    /K     1
    /Lang    (EN-US)
  ⊞ /P     <<Dictionary>>
  ⊞ /Pg    <<Dictionary>>
    /S    /H1
```

The Portable Document Fo

Imaging model

The basic design of how graphics are represented in PDF is v
except for the use of transparency, which was added in PDF

matrix (CTM), which determines

which is a key component of trar

Image XObjects) are represente

tension	Description
g	a lossy filter based
f	a lossless bi-le Group 3 or Gr standard defi

age

by text elements in page content
rawn at certain positions. The cha
resource.

scription of a digital typeface. It r
or it may include an embedded

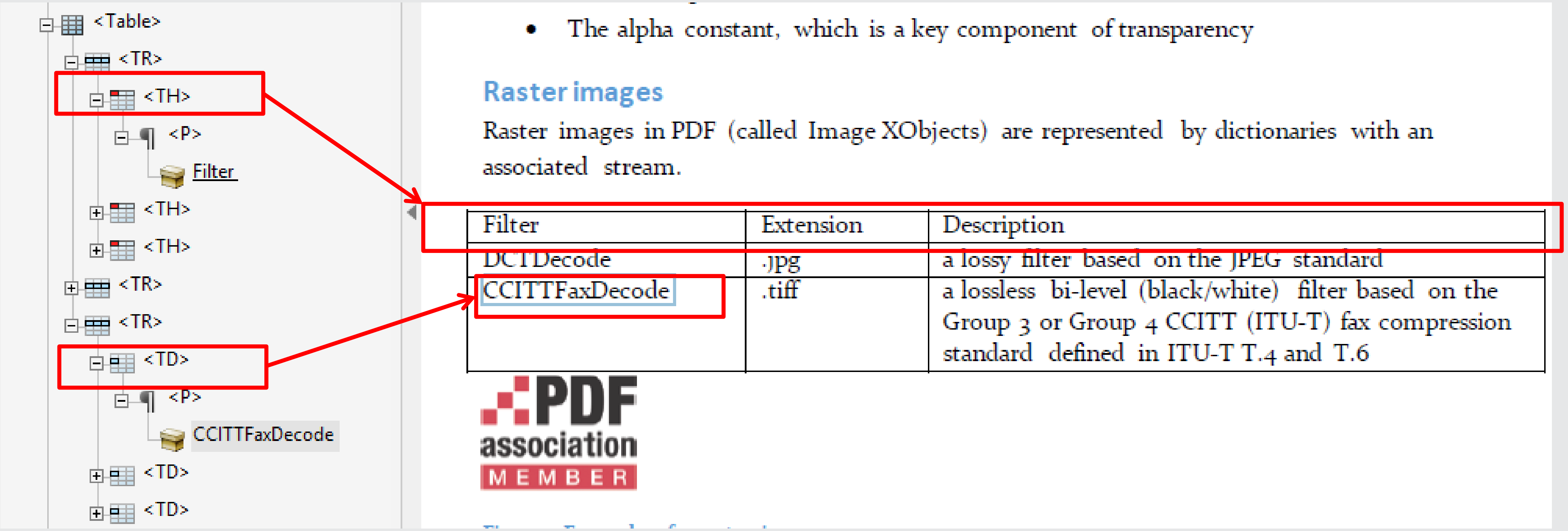
```
<h1 lang="EN-US">
    <span>Imaging model</span>
</h1>
```

Imaging model

The basic design of how graphics are represented in PDF is very similar to that of PostScript, exc

The current transformation matrix (CTM), which determines the coordinate system

Table



```
<Table>
  <TR>
    <TH>
      <P>
        Filter
    <TH>
    <TH>
  <TR>
  <TR>
    <TD>
      <P>
        CCITTFaxDecode
    <TD>
    <TD>
```

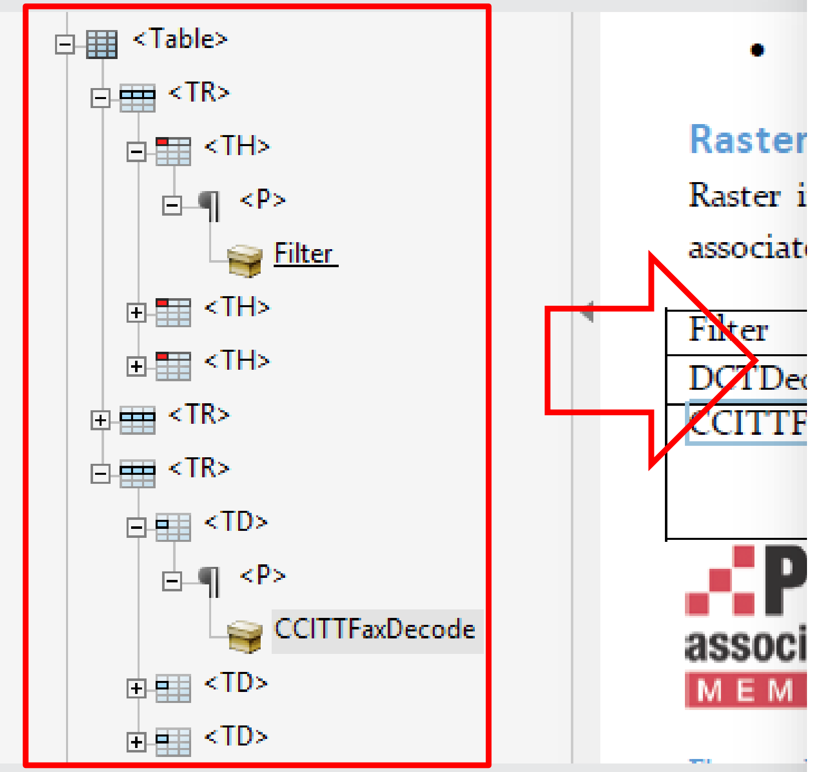
- The alpha constant, which is a key component of transparency

Raster images

Raster images in PDF (called Image XObjects) are represented by dictionaries with an associated stream.

Filter	Extension	Description
DCTDecode	.jpg	a lossy filter based on the JPEG standard
CCITTFaxDecode	.tiff	a lossless bi-level (black/white) filter based on the Group 3 or Group 4 CCITT (ITU-T) fax compression standard defined in ITU-T T.4 and T.6

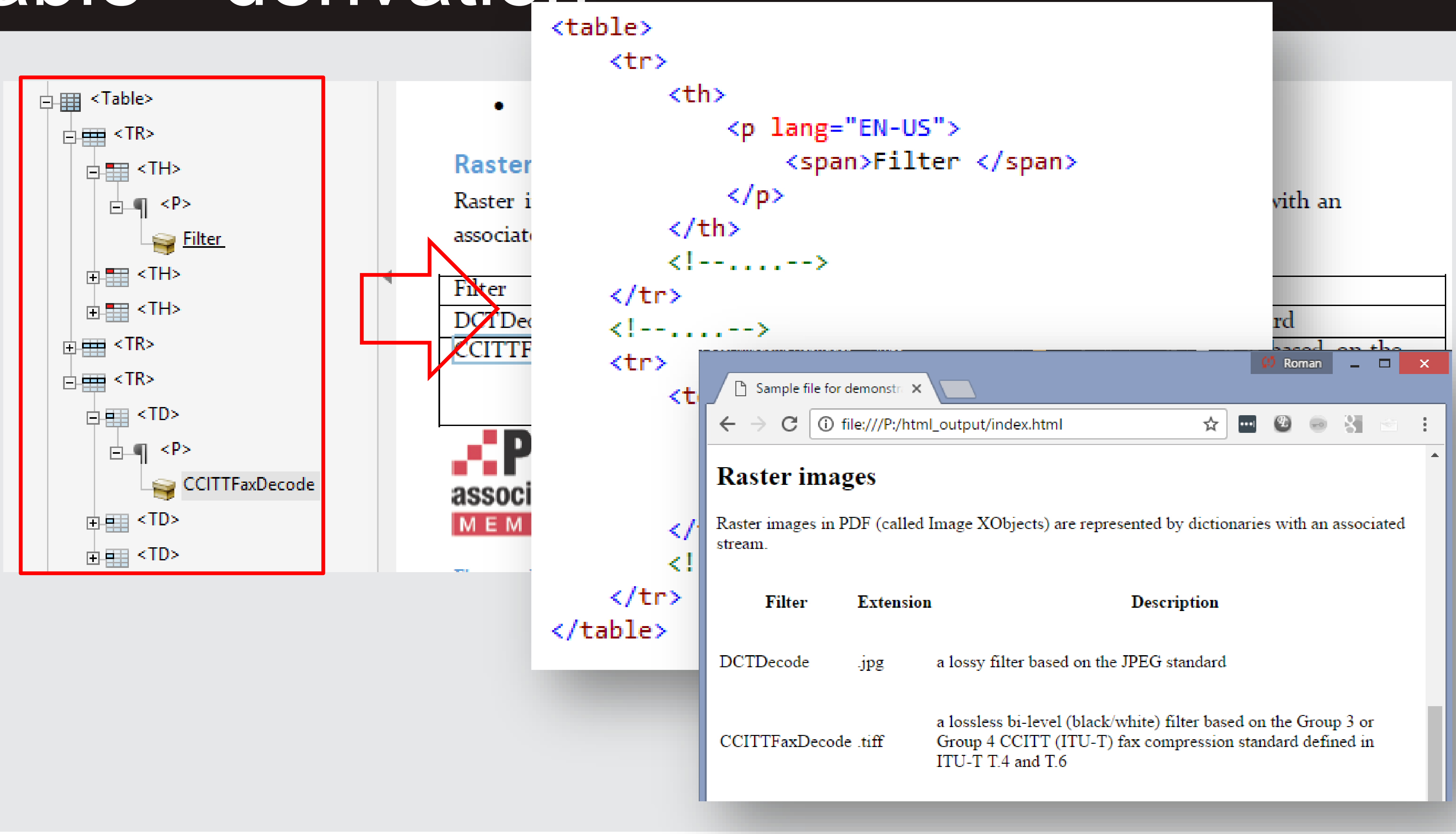
Table - derivation



```
<Table>
  <TR>
    <TH>
      <P>
        Filter
    <TH>
    <TH>
  <TR>
  <TR>
    <TD>
      <P>
        CCITTFaxDecode
    <TD>
    <TD>
```

```
<table>
    <tr>
        <th>
            <p lang="EN-US">
                <span>Filter </span>
            </p>
        </th>
        <!--.....-->
    </tr>
    <!--.....-->
    <tr>
        <td>
            <p lang="EN-US">
                <span>CCITTFaxDecode </span>
            </p>
        </td>
        <!--.....-->
    </tr>
</table>
```

Table - derivation



```
<Table>
    <TR>
        <TH>
            <P>
                Filter
        <TH>
        <TH>
    <TR>
    <TR>
        <TD>
            <P>
                CCITTFaxDecode
        <TD>
        <TD>
```

```
<table>
    <tr>
        <th>
            <p lang="EN-US">
                <span>Filter </span>
            </p>
        </th>
        <!--.....-->
    </tr>
    <!--.....-->
    <tr>
        <t

        </
        <!
    </tr>
</table>
```

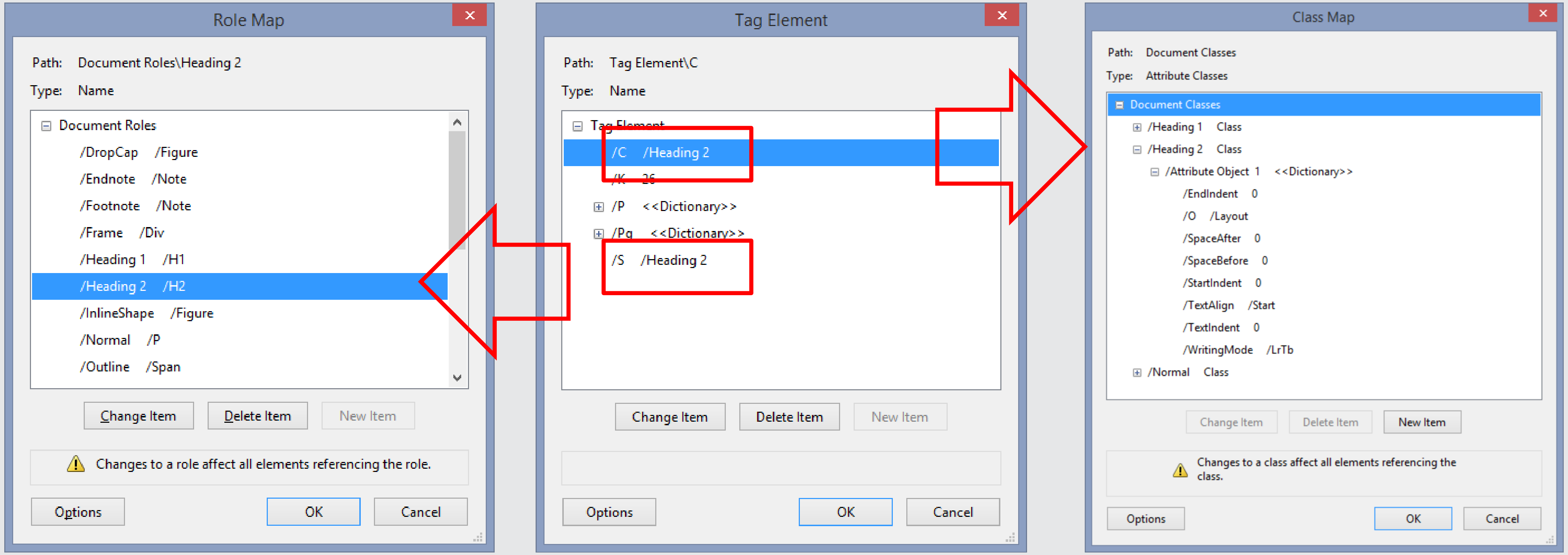
Raster images

Raster images in PDF (called Image XObjects) are represented by dictionaries with an associated stream.

Filter	Extension	Description
DCTDecode	.jpg	a lossy filter based on the JPEG standard
CCITTFaxDecode	.tiff	a lossless bi-level (black/white) filter based on the Group 3 or Group 4 CCITT (ITU-T) fax compression standard defined in ITU-T T.4 and T.6

Attributes – general concept

- Each tag could have attributes
- Different practice is using ClassMap
- Role mapping takes place



Path: Document Roles\Heading 2

Type: Name

```
⊟ Document Roles
    /DropCap     /Figure
    /Endnote     /Note
    /Footnote    /Note
    /Frame    /Div
    /Heading 1    /H1
    /Heading 2    /H2
    /InlineShape    /Figure
    /Normal    /P
    /Outline    /Span
```

⚠ Changes to a role affect all elements referencing the role.

Path: Tag Element\C

Type: Name

```
⊟ Tag Element
    /C    /Heading 2
    /K    26
  ⊞ /P    <<Dictionary>>
  ⊞ /Pg    <<Dictionary>>
    /S    /Heading 2
```

Path: Document Classes

Type: Attribute Classes

```
⊟ Document Classes
  ⊞ /Heading 1    Class
  ⊟ /Heading 2    Class
    ⊟ /Attribute Object 1    <<Dictionary>>
        /EndIndent    0
        /O    /Layout
        /SpaceAfter    0
        /SpaceBefore    0
        /StartIndent    0
        /TextAlign    /Start
        /TextIndent    0
        /WritingMode    /LrTb
  ⊞ /Normal    Class
```

⚠ Changes to a class affect all elements referencing the class.

Attribute owners

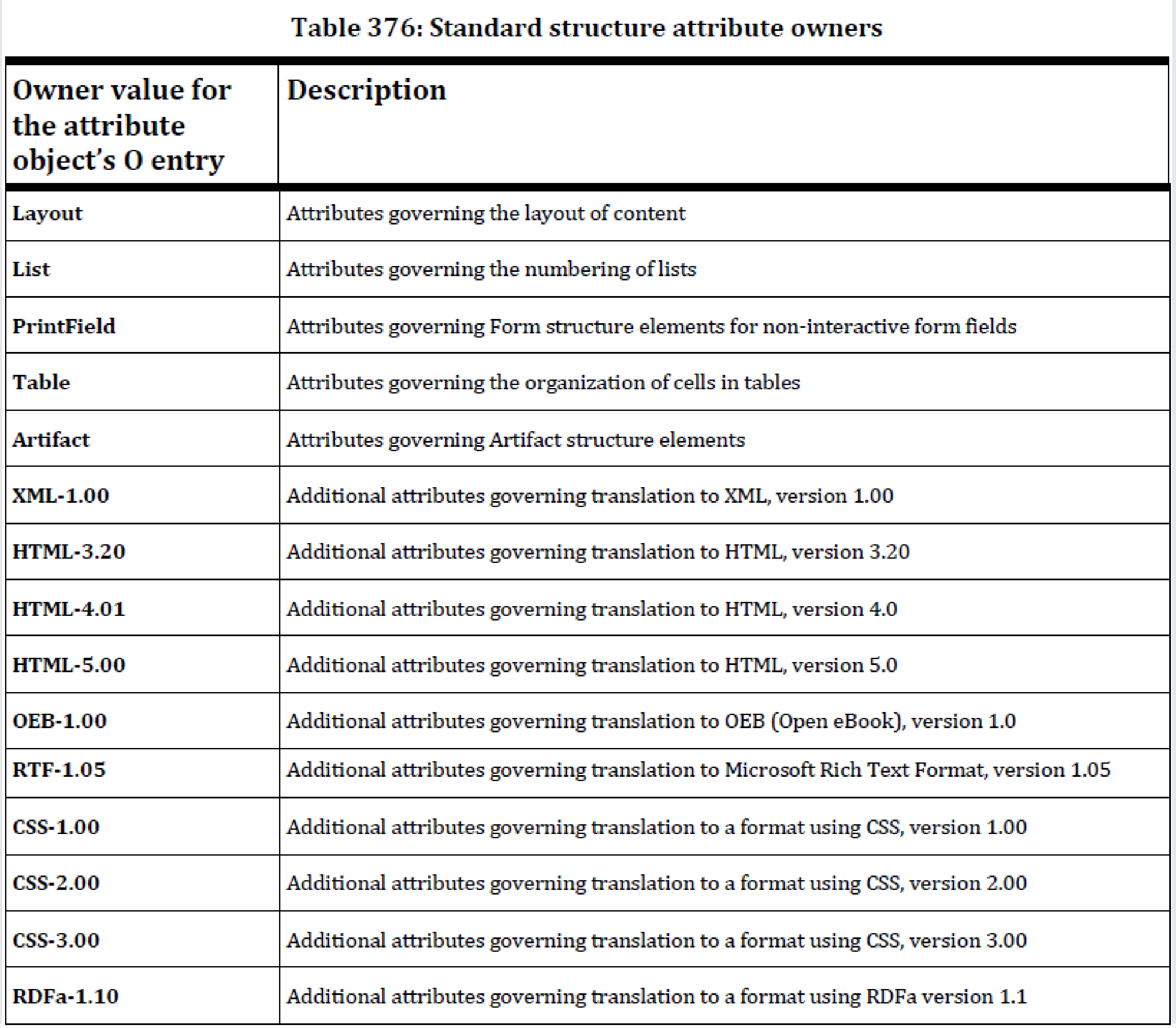


Table 376: Standard structure attribute owners

Owner value for the attribute object's O entry	Description
Layout	Attributes governing the layout of content
List	Attributes governing the numbering of lists
PrintField	Attributes governing Form structure elements for non-interactive form fields
Table	Attributes governing the organization of cells in tables
Artifact	Attributes governing Artifact structure elements
XML-1.00	Additional attributes governing translation to XML, version 1.00
HTML-3.20	Additional attributes governing translation to HTML, version 3.20
HTML-4.01	Additional attributes governing translation to HTML, version 4.0
HTML-5.00	Additional attributes governing translation to HTML, version 5.0
OEB-1.00	Additional attributes governing translation to OEB (Open eBook), version 1.0
RTF-1.05	Additional attributes governing translation to Microsoft Rich Text Format, version 1.05
CSS-1.00	Additional attributes governing translation to a format using CSS, version 1.00
CSS-2.00	Additional attributes governing translation to a format using CSS, version 2.00
CSS-3.00	Additional attributes governing translation to a format using CSS, version 3.00
RDFa-1.10	Additional attributes governing translation to a format using RDFa version 1.1

Attribute owners

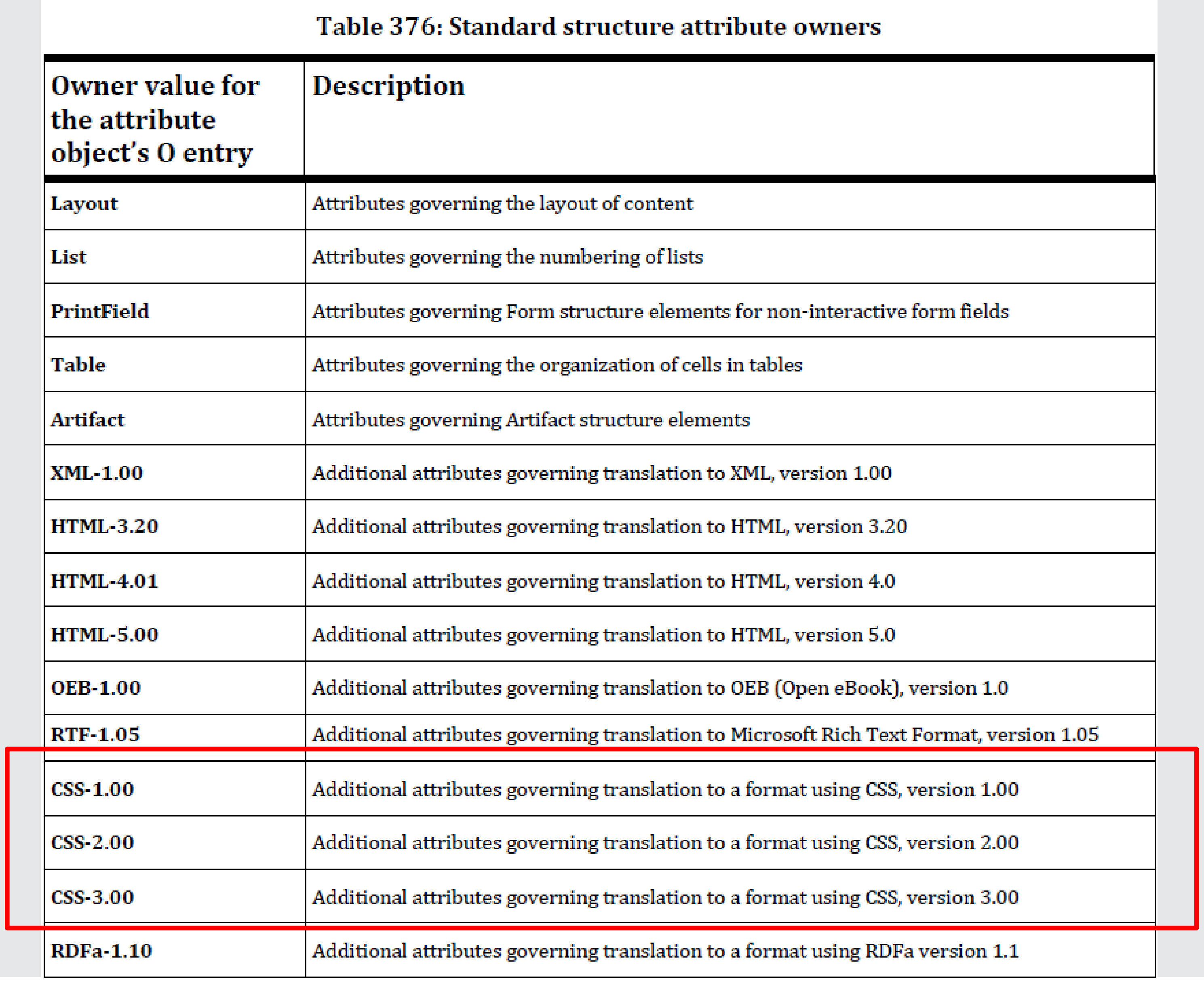
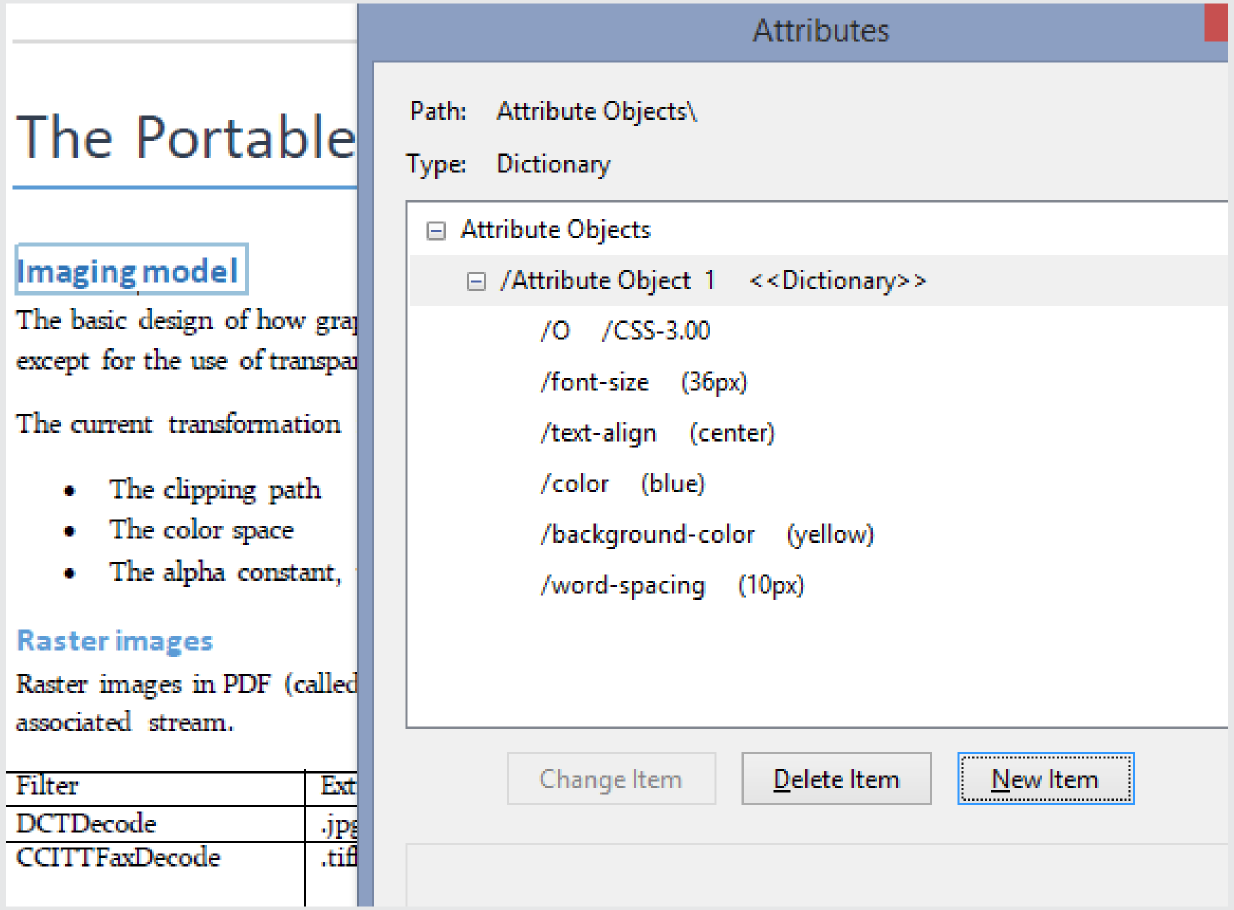


Table 376: Standard structure attribute owners

Owner value for the attribute object's O entry	Description
Layout	Attributes governing the layout of content
List	Attributes governing the numbering of lists
PrintField	Attributes governing Form structure elements for non-interactive form fields
Table	Attributes governing the organization of cells in tables
Artifact	Attributes governing Artifact structure elements
XML-1.00	Additional attributes governing translation to XML, version 1.00
HTML-3.20	Additional attributes governing translation to HTML, version 3.20
HTML-4.01	Additional attributes governing translation to HTML, version 4.0
HTML-5.00	Additional attributes governing translation to HTML, version 5.0
OEB-1.00	Additional attributes governing translation to OEB (Open eBook), version 1.0
RTF-1.05	Additional attributes governing translation to Microsoft Rich Text Format, version 1.05
CSS-1.00	Additional attributes governing translation to a format using CSS, version 1.00
CSS-2.00	Additional attributes governing translation to a format using CSS, version 2.00
CSS-3.00	Additional attributes governing translation to a format using CSS, version 3.00
RDFa-1.10	Additional attributes governing translation to a format using RDFa version 1.1

Attribute usage



The Portable

Imaging model

The basic design of how grap
except for the use of transpa

The current transformation

- The clipping path
- The color space
- The alpha constant,

Raster images

Raster images in PDF (called
associated stream.

Filter	Ext
DCTDecode	.jpg
CCITTFaxDecode	.tif

Attributes

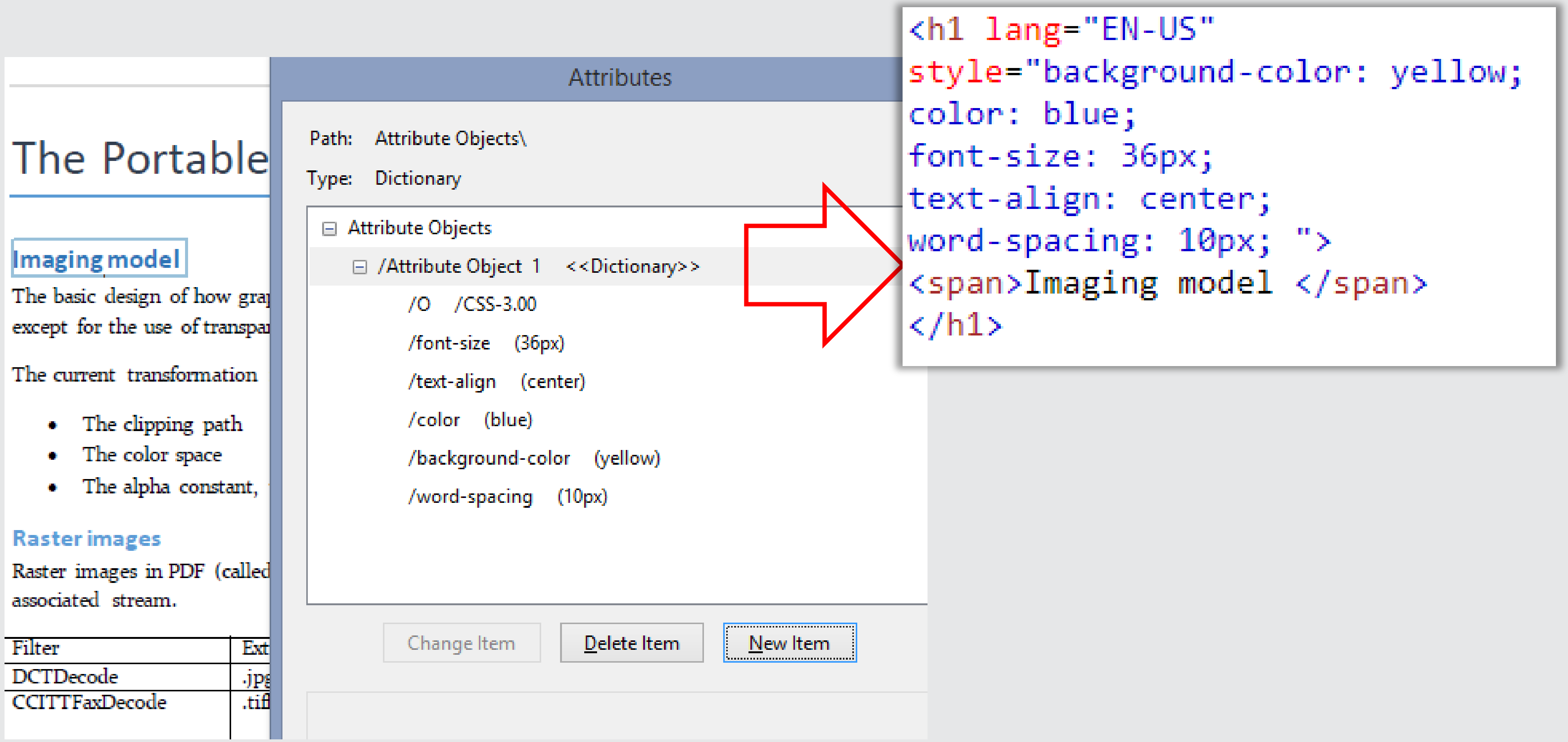
Path: Attribute Objects\

Type: Dictionary

```
⊟ Attribute Objects
   ⊟ /Attribute Object 1    <<Dictionary>>
        /O    /CSS-3.00
        /font-size    (36px)
        /text-align    (center)
        /color    (blue)
        /background-color    (yellow)
        /word-spacing    (10px)
```

- Layout
- Table
- List
- Can use the power of CSS

Attribute usage



The Portable

Imaging model

The basic design of how gra

except for the use of transpa

The current transformation

- The clipping path
- The color space
- The alpha constant,

Raster images

Raster images in PDF (called

associated stream.

Filter	Ext
DCTDecode	.jpg
CCITTFaxDecode	.tif

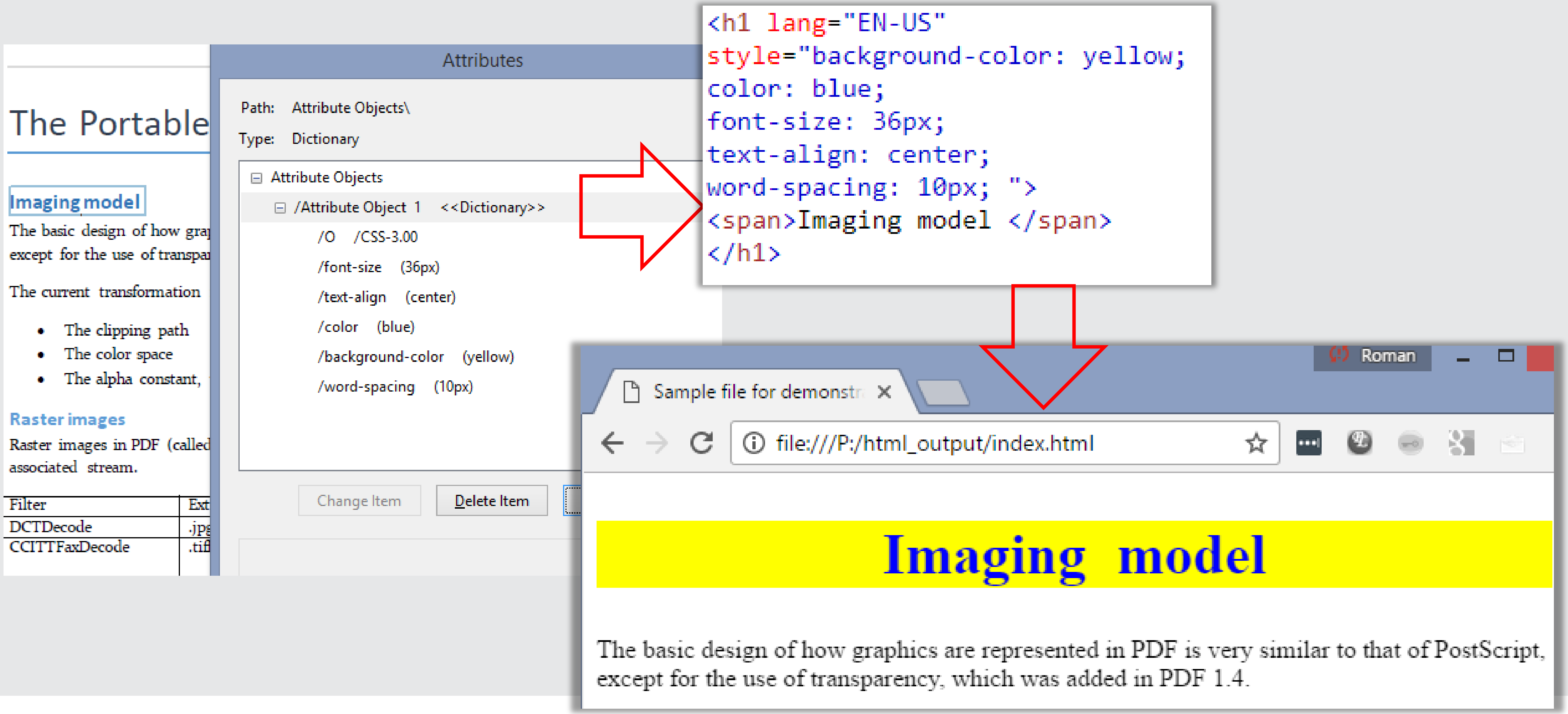
Path: Attribute Objects\

Type: Dictionary

```
⊟ Attribute Objects
  ⊟ /Attribute Object 1    <<Dictionary>>
      /O    /CSS-3.00
      /font-size    (36px)
      /text-align    (center)
      /color    (blue)
      /background-color    (yellow)
      /word-spacing    (10px)
```

```html
<h1 lang="EN-US"
style="background-color: yellow;
color: blue;
font-size: 36px;
text-align: center;
word-spacing: 10px; ">
<span>Imaging model </span>
</h1>
```

Attribute usage



The Portable

Imaging model

The basic design of how gra
except for the use of transpa

The current transformation

- The clipping path
- The color space
- The alpha constant,

Raster images

Raster images in PDF (called
associated stream.

Filter	Ext
DCTDecode	.jpg
CCITTFaxDecode	.tif

Attributes

Path: Attribute Objects\

Type: Dictionary

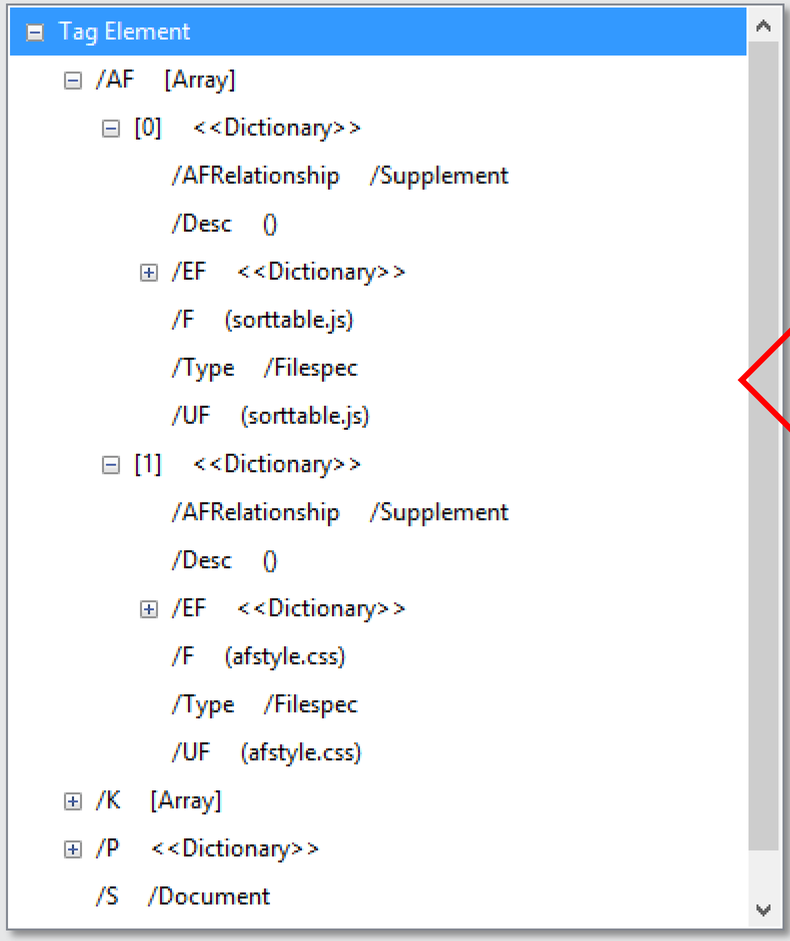
```
⊟ Attribute Objects
  ⊟ /Attribute Object 1   <<Dictionary>>
        /O   /CSS-3.00
        /font-size   (36px)
        /text-align   (center)
        /color   (blue)
        /background-color   (yellow)
        /word-spacing   (10px)
```

```
<h1 lang="EN-US"
style="background-color: yellow;
color: blue;
font-size: 36px;
text-align: center;
word-spacing: 10px; ">
<span>Imaging model </span>
</h1>
```

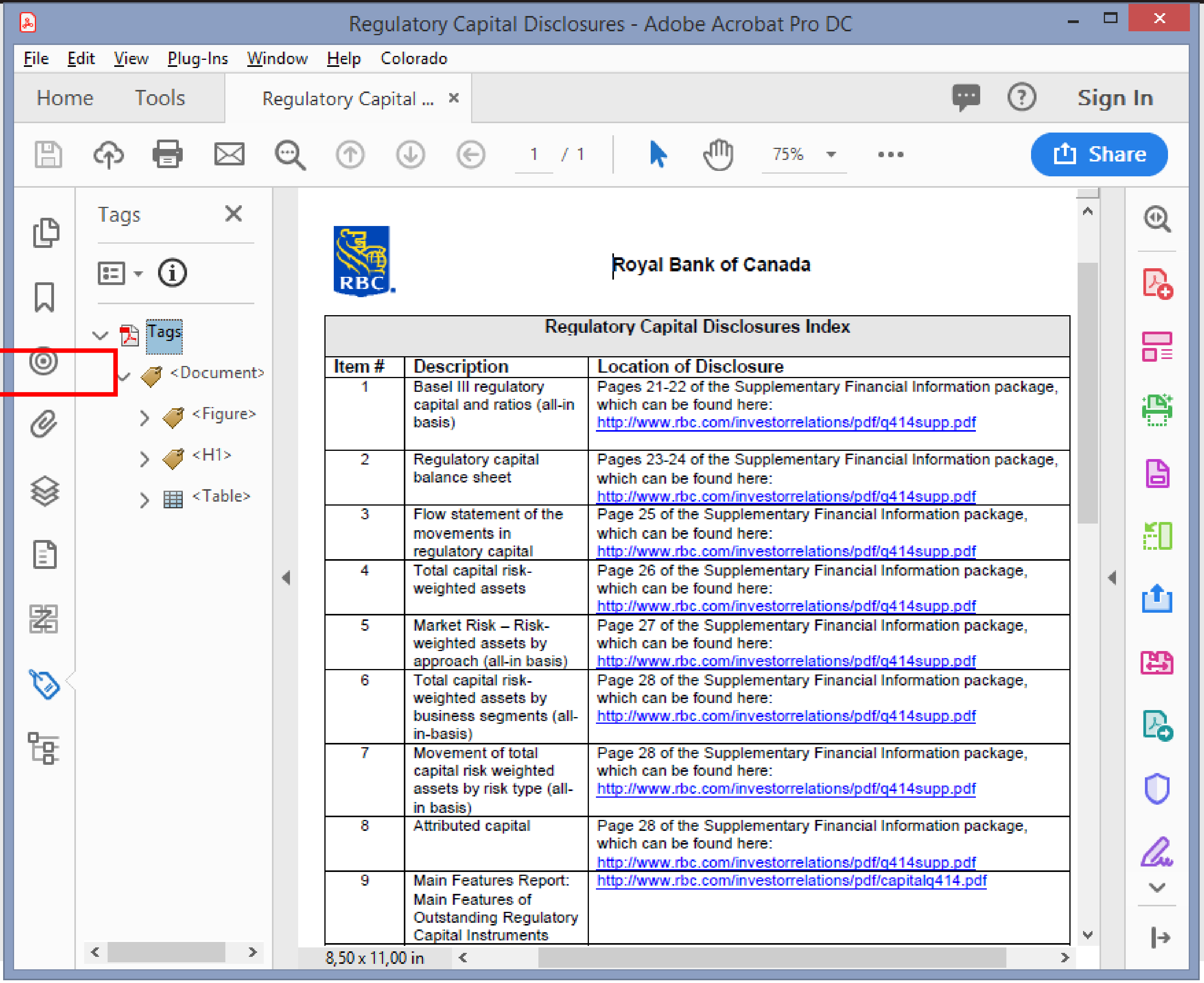
Imaging model

The basic design of how graphics are represented in PDF is very similar to that of PostScript, except for the use of transparency, which was added in PDF 1.4.

Associated files – css, js, svg, mathml



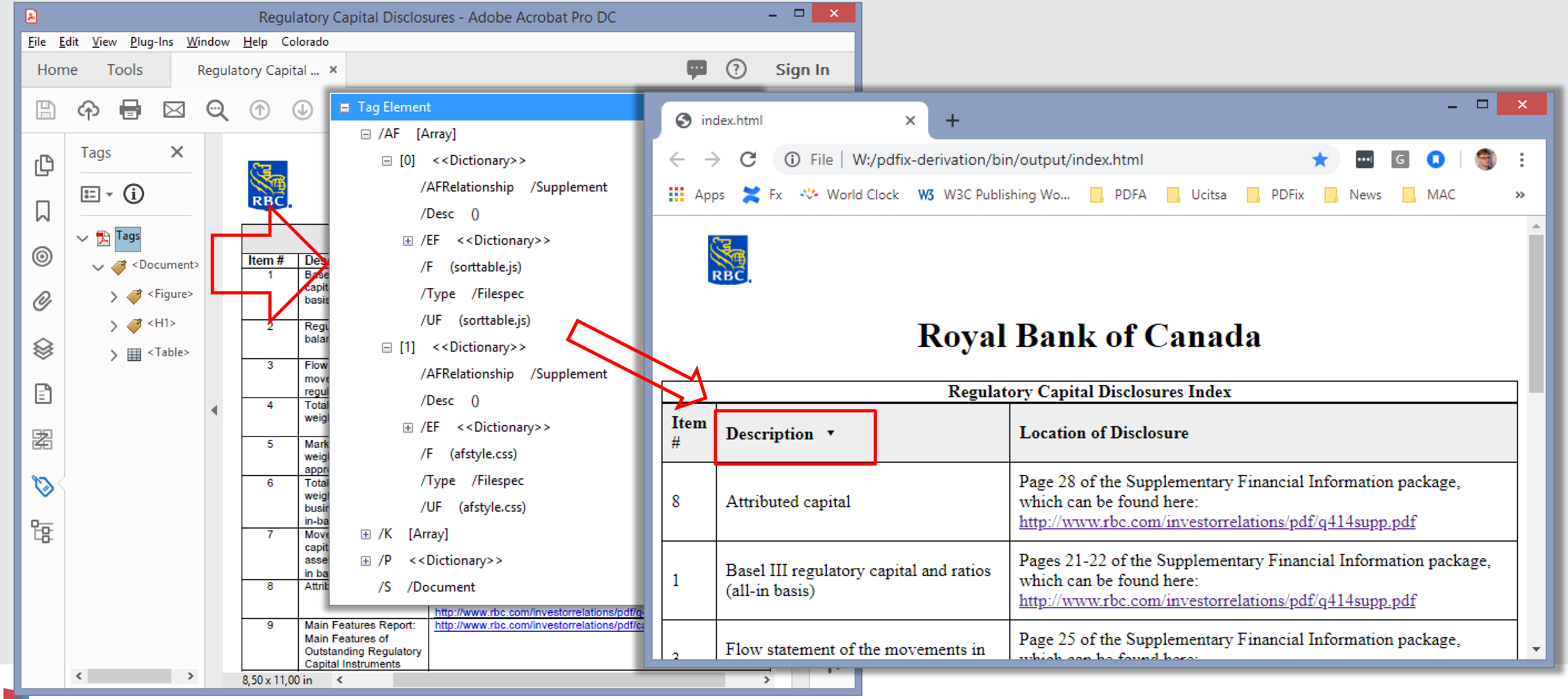
```
⊟ Tag Element
  ⊟ /AF   [Array]
    ⊟ [0]   <<Dictionary>>
        /AFRelationship   /Supplement
        /Desc    ()
      ⊞ /EF    <<Dictionary>>
        /F    (sorttable.js)
        /Type    /Filespec
        /UF    (sorttable.js)
    ⊟ [1]   <<Dictionary>>
        /AFRelationship   /Supplement
        /Desc    ()
      ⊞ /EF    <<Dictionary>>
        /F    (afstyle.css)
        /Type    /Filespec
        /UF    (afstyle.css)
  ⊞ /K    [Array]
  ⊞ /P    <<Dictionary>>
    /S    /Document
```


Regulatory Capital Disclosures - Adobe Acrobat Pro DC
File Edit View Plug-Ins Window Help Colorado
Home Tools Regulatory Capital ... ×
Sign In
1 / 1
75%
Share
Tags
Tags
<Document>
<Figure>
<H1>
<Table>
8,50 x 11,00 in

Royal Bank of Canada

Regulatory Capital Disclosures Index		
Item #	Description	Location of Disclosure
1	Basel III regulatory capital and ratios (all-in basis)	Pages 21-22 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
2	Regulatory capital balance sheet	Pages 23-24 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
3	Flow statement of the movements in regulatory capital	Page 25 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
4	Total capital risk-weighted assets	Page 26 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
5	Market Risk – Risk-weighted assets by approach (all-in basis)	Page 27 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
6	Total capital risk-weighted assets by business segments (all-in-basis)	Page 28 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
7	Movement of total capital risk weighted assets by risk type (all-in basis)	Page 28 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
8	Attributed capital	Page 28 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
9	Main Features Report: Main Features of Outstanding Regulatory Capital Instruments	http://www.rbc.com/investorrelations/pdf/capitalq414.pdf

Associated files – css, js, svg, mathml



Regulatory Capital Disclosures - Adobe Acrobat Pro DC
File Edit View Plug-Ins Window Help Colorado
Home Tools Regulatory Capital ... ×
Sign In
Tags
Tags
<Document>
<Figure>
<H1>
<Table>
RBC
Item #
8,50 x 11,00 in
index.html
File W:/pdfix-derivation/bin/output/index.html
Apps Fx World Clock W3C Publishing Wo... PDFA Ucitsa PDFix News MAC

```
Tag Element
  /AF    [Array]
    [0]    <<Dictionary>>
        /AFRelationship    /Supplement
        /Desc    ()
      /EF    <<Dictionary>>
        /F    (sorttable.js)
        /Type    /Filespec
        /UF    (sorttable.js)
    [1]    <<Dictionary>>
        /AFRelationship    /Supplement
        /Desc    ()
      /EF    <<Dictionary>>
        /F    (afstyle.css)
        /Type    /Filespec
        /UF    (afstyle.css)
  /K    [Array]
  /P    <<Dictionary>>
  /S    /Document
```

Item #	Des
1	Base capit basis
2	Regu balar
3	Flow move regul
4	Total weig
5	Mark weig appr
6	Total weig busir in-ba
7	Move capit asse in ba
8	Attrib
9	Main Features Report: Main Features of Outstanding Regulatory Capital Instruments

http://www.rbc.com/investorrelations/pdf/q
http://www.rbc.com/investorrelations/pdf/ca

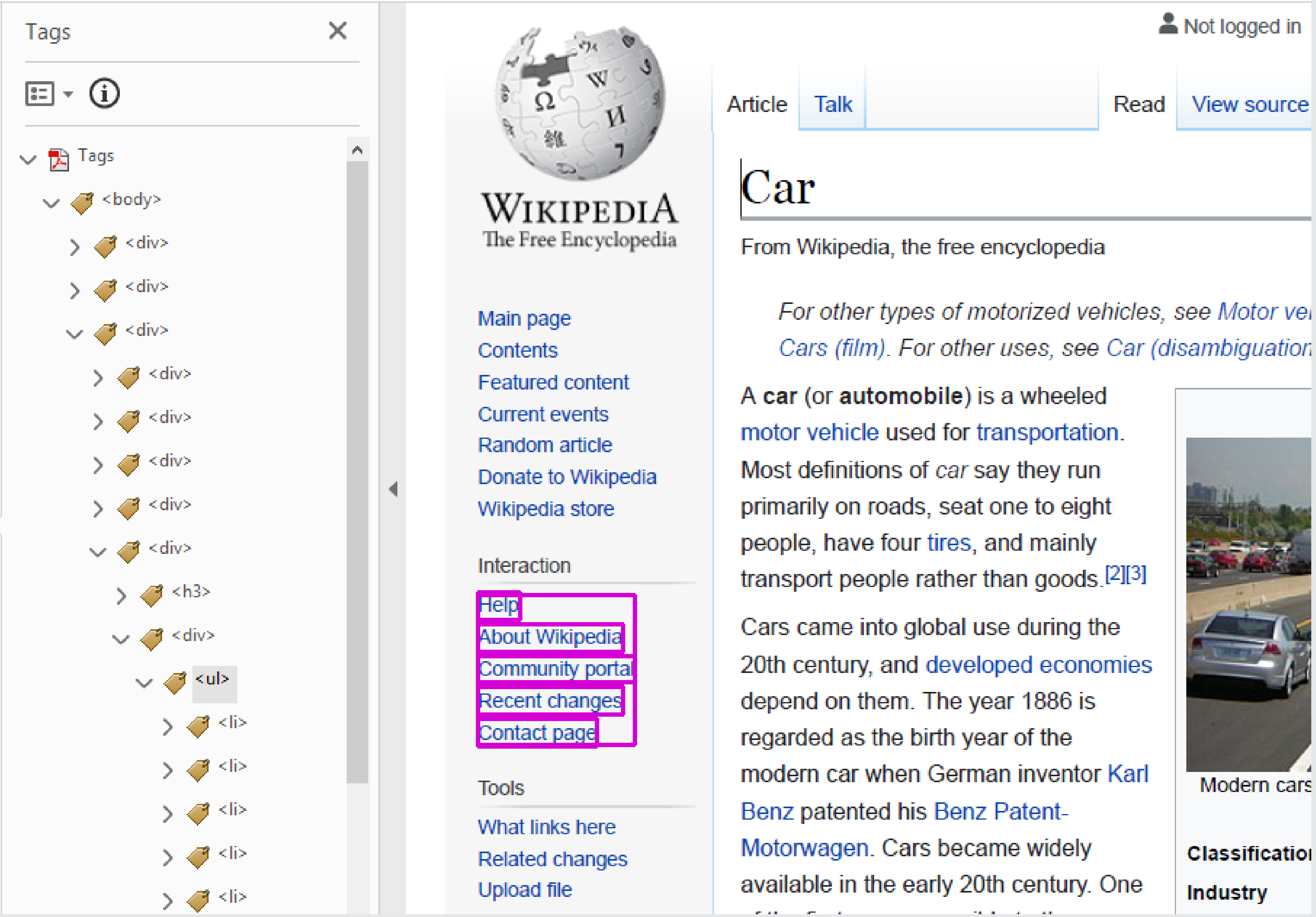
Royal Bank of Canada

Regulatory Capital Disclosures Index

Item #	Description ▼	Location of Disclosure
8	Attributed capital	Page 28 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
1	Basel III regulatory capital and ratios (all-in basis)	Pages 21-22 of the Supplementary Financial Information package, which can be found here: http://www.rbc.com/investorrelations/pdf/q414supp.pdf
3	Flow statement of the movements in	Page 25 of the Supplementary Financial Information package, which can be found here:

Roman Toda NORMEX

Namespaces – sample with html



Tags

Tags
- <body>
  - <div>
  - <div>
  - <div>
    - <div>
    - <div>
    - <div>
    - <div>
    - <div>
      - <h3>
      - <div>
        - <ul>
          - <li>
          - <li>
          - <li>
          - <li>
          - <li>

Not logged in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read View source

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file

Car

From Wikipedia, the free encyclopedia

For other types of motorized vehicles, see Motor ve
Cars (film). For other uses, see Car (disambiguation

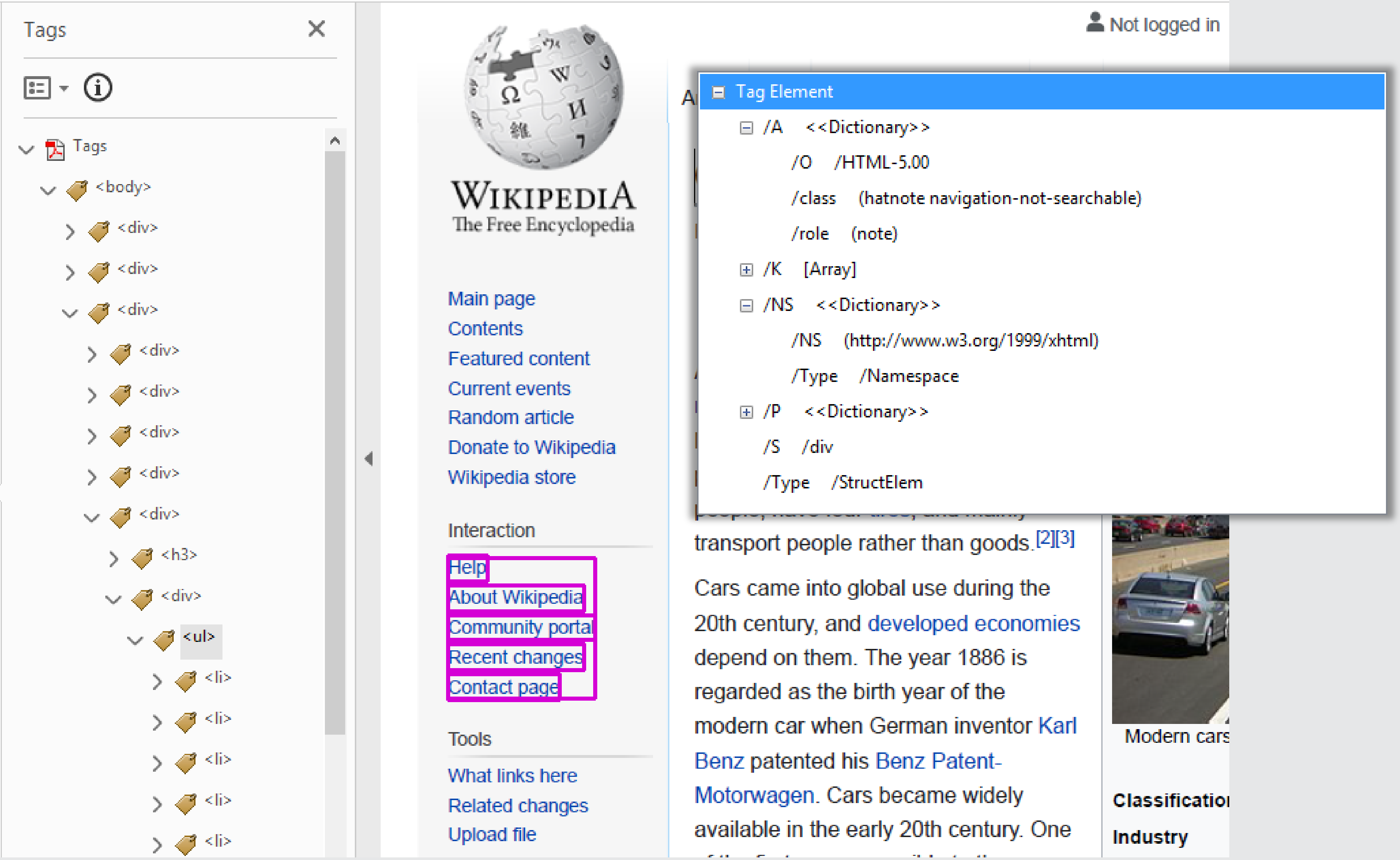
A **car** (or **automobile**) is a wheeled motor vehicle used for transportation. Most definitions of *car* say they run primarily on roads, seat one to eight people, have four tires, and mainly transport people rather than goods.[2][3]

Cars came into global use during the 20th century, and developed economies depend on them. The year 1886 is regarded as the birth year of the modern car when German inventor Karl Benz patented his Benz Patent-Motorwagen. Cars became widely available in the early 20th century. One

Modern cars

Classificatio
Industry

Namespaces – sample with html



```
Tags
  Tags
    <body>
      <div>
      <div>
      <div>
        <div>
        <div>
        <div>
        <div>
        <div>
          <h3>
          <div>
            <ul>
              <li>
              <li>
              <li>
              <li>
              <li>
```

```
Tag Element
  /A    <<Dictionary>>
      /O    /HTML-5.00
      /class    (hatnote navigation-not-searchable)
      /role    (note)
  /K    [Array]
  /NS    <<Dictionary>>
      /NS    (http://www.w3.org/1999/xhtml)
      /Type    /Namespace
  /P    <<Dictionary>>
  /S    /div
  /Type    /StructElem
```

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file

transport people rather than goods.[2][3]

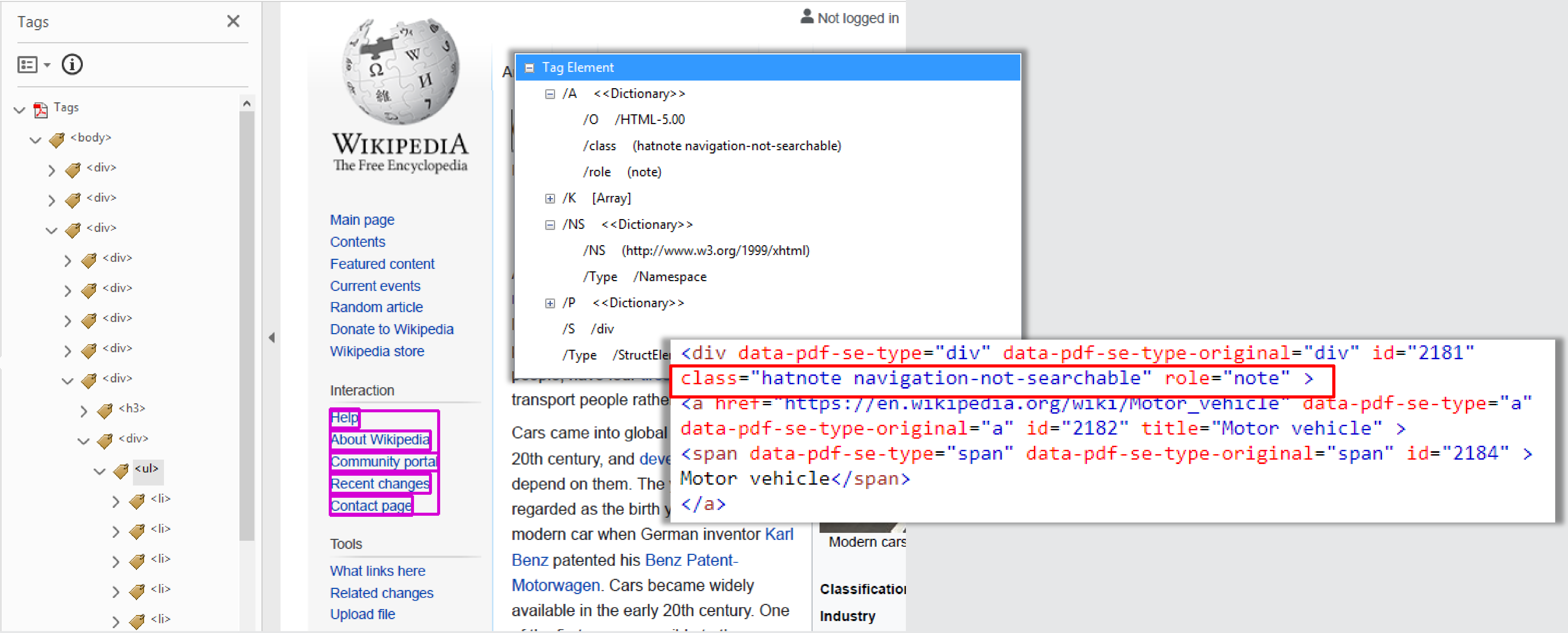
Cars came into global use during the 20th century, and developed economies depend on them. The year 1886 is regarded as the birth year of the modern car when German inventor Karl Benz patented his Benz Patent-Motorwagen. Cars became widely available in the early 20th century. One

Modern cars

Classificatio

Industry

Namespaces – sample with html



Tags

```
Tags
  <body>
    <div>
    <div>
    <div>
      <div>
      <div>
      <div>
      <div>
      <div>
        <h3>
        <div>
          <ul>
            <li>
            <li>
            <li>
            <li>
            <li>
```

Not logged in

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file

```
Tag Element
  /A    <<Dictionary>>
      /O     /HTML-5.00
      /class    (hatnote navigation-not-searchable)
      /role    (note)
  /K    [Array]
  /NS    <<Dictionary>>
      /NS    (http://www.w3.org/1999/xhtml)
      /Type    /Namespace
  /P    <<Dictionary>>
  /S    /div
  /Type    /StructEle
```

```
<div data-pdf-se-type="div" data-pdf-se-type-original="div" id="2181"
class="hatnote navigation-not-searchable" role="note" >
<a href="https://en.wikipedia.org/wiki/Motor_vehicle" data-pdf-se-type="a"
data-pdf-se-type-original="a" id="2182" title="Motor vehicle" >
<span data-pdf-se-type="span" data-pdf-se-type-original="span" id="2184" >
Motor vehicle</span>
</a>
```

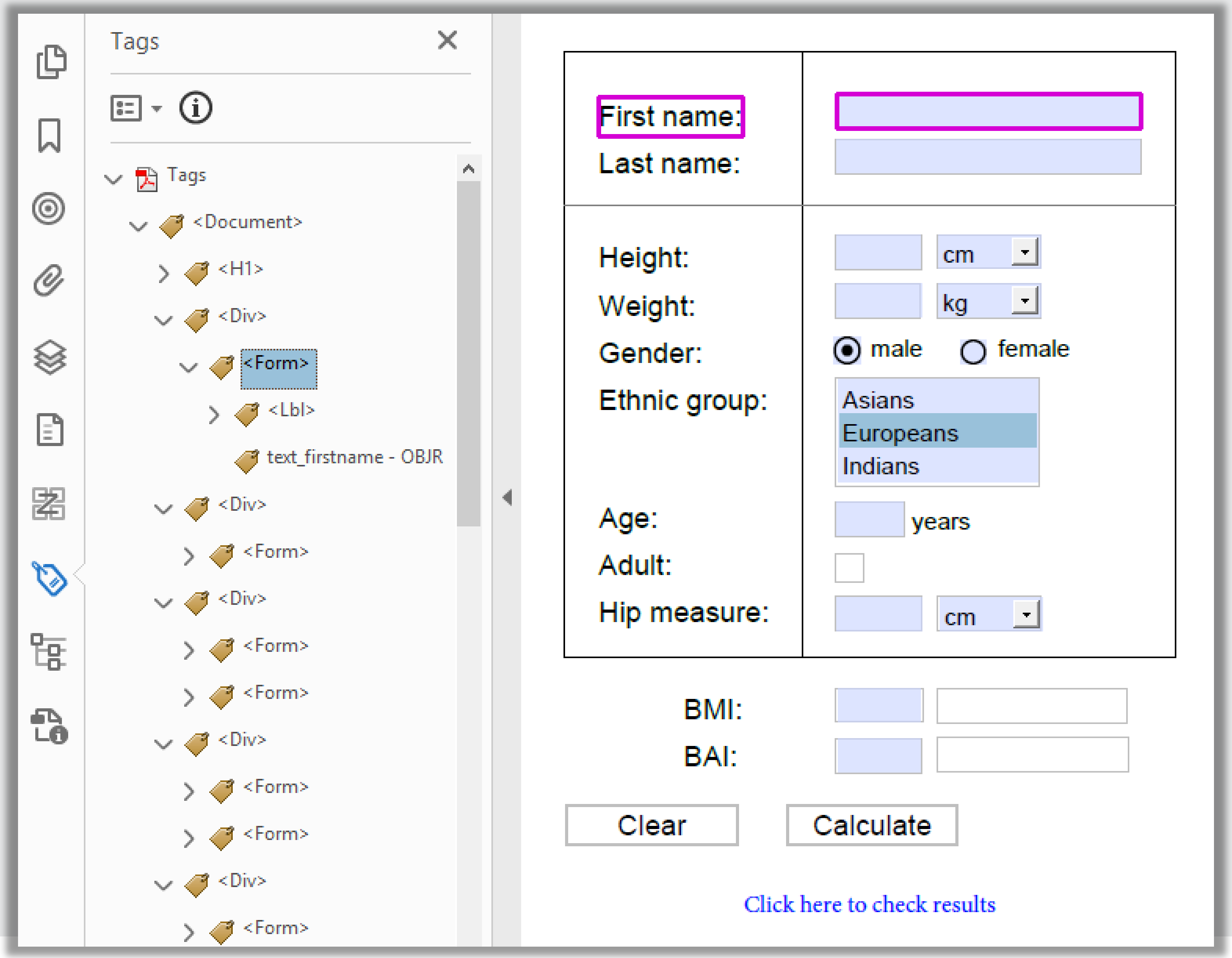
transport people rathe

Cars came into global
20th century, and deve
depend on them. The
regarded as the birth y
modern car when German inventor Karl
Benz patented his Benz Patent-
Motorwagen. Cars became widely
available in the early 20th century. One

Modern cars

Classificatio
Industry

Forms sample



Tags

- Tags
  - <Document>
    - <H1>
    - <Div>
      - <Form>
        - <Lbl>
        - text_firstname - OBJR
    - <Div>
      - <Form>
    - <Div>
      - <Form>
      - <Form>
    - <Div>
      - <Form>
      - <Form>
    - <Div>
      - <Form>

First name:

Last name:

Height: cm

Weight: kg

Gender: ● male ○ female

Ethnic group: Asians, Europeans, Indians

Age: years

Adult: □

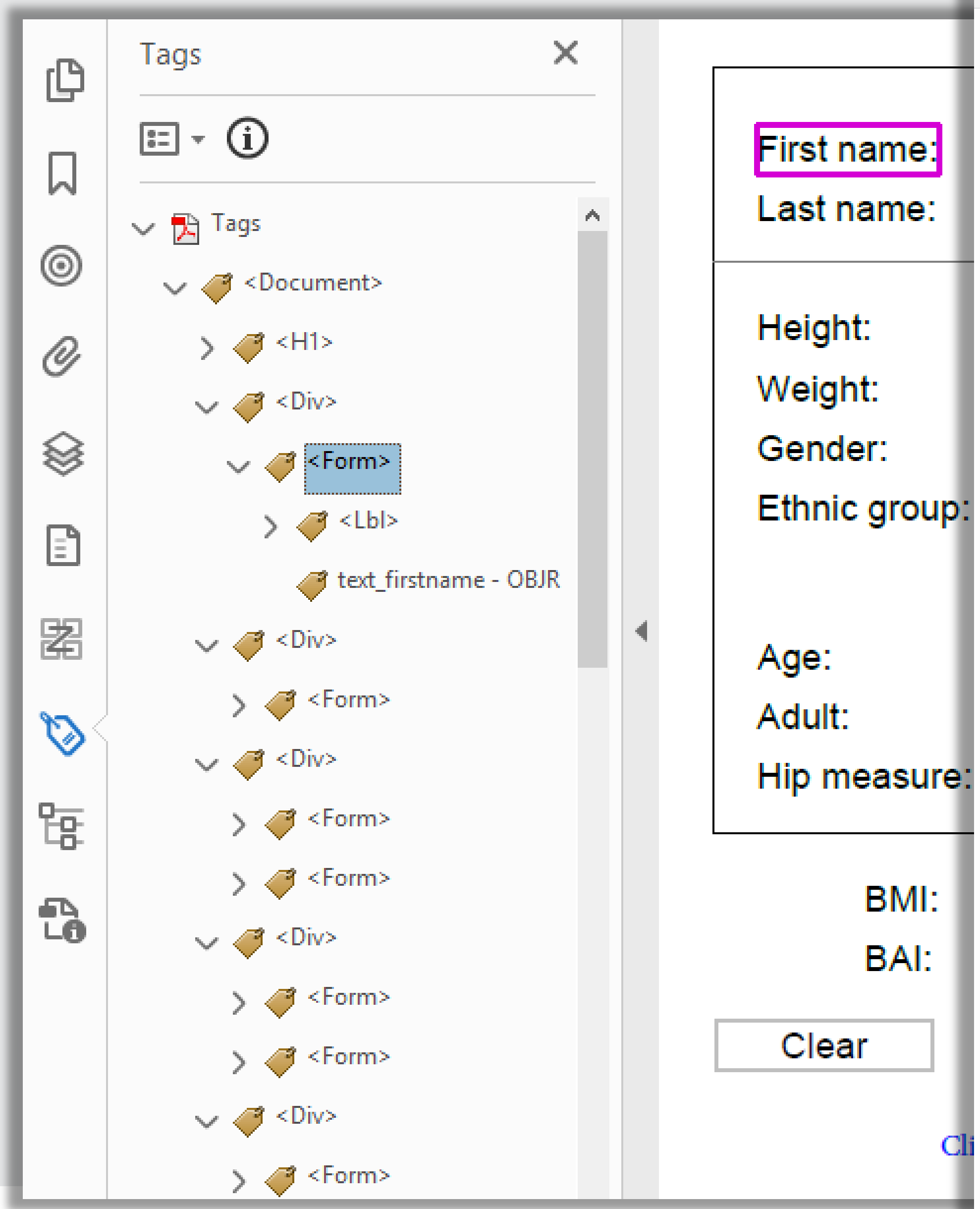
Hip measure: cm

BMI:

BAI:

Click here to check results

Forms sample



Tags

- <Document>
  - <H1>
  - <Div>
    - <Form>
      - <Lbl>
      - text_firstname - OBJR
  - <Div>
    - <Form>
  - <Div>
    - <Form>
    - <Form>
  - <Div>
    - <Form>
    - <Form>
  - <Div>
    - <Form>

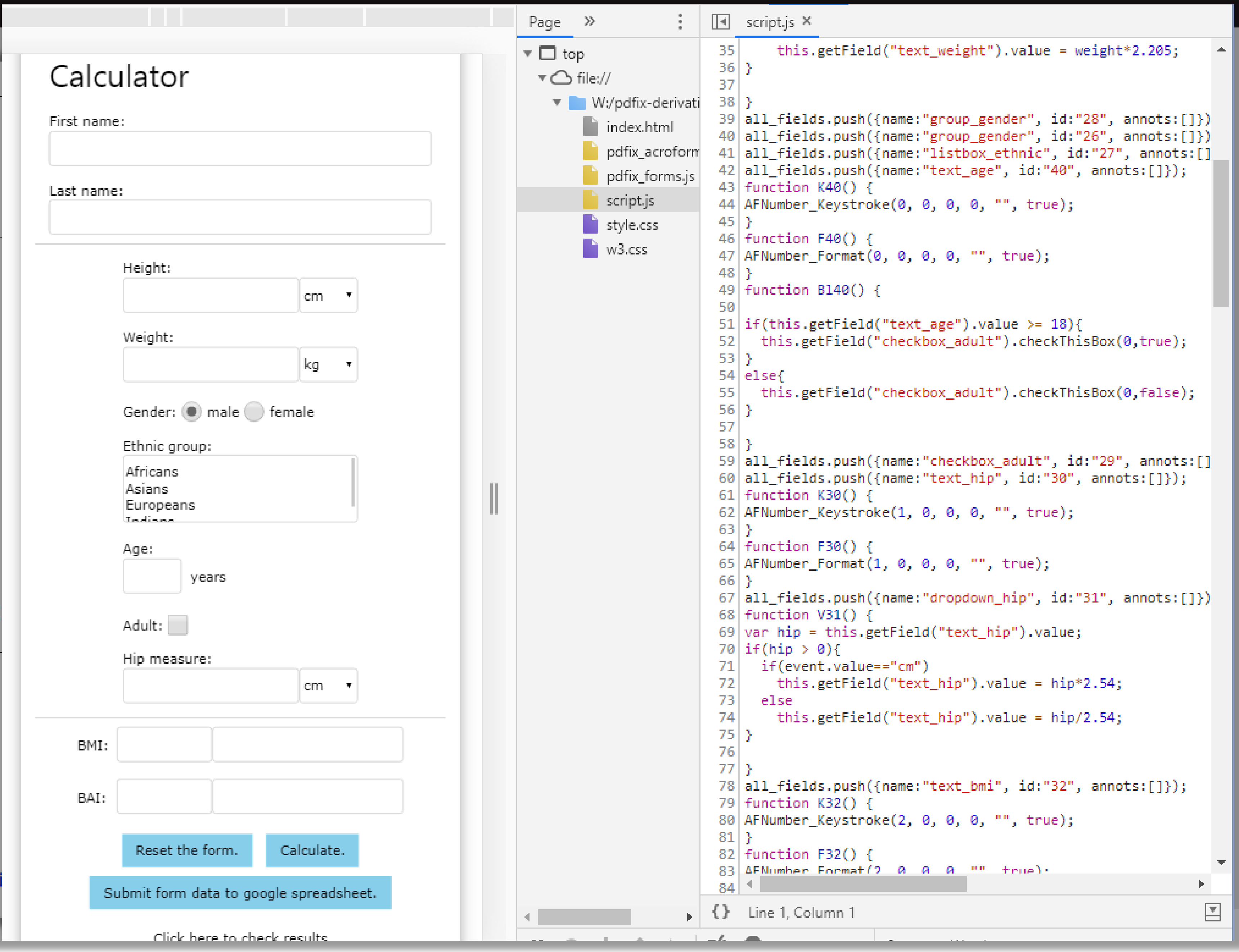
First name:
Last name:

Height:
Weight:
Gender:
Ethnic group:

Age:
Adult:
Hip measure:

BMI:
BAI:

Clear



Calculator

First name:

Last name:

Height: cm

Weight: kg

Gender: ● male ○ female

Ethnic group:
Africans
Asians
Europeans
Indians

Age: years

Adult: □

Hip measure: cm

BMI:

BAI:

Reset the form. Calculate.

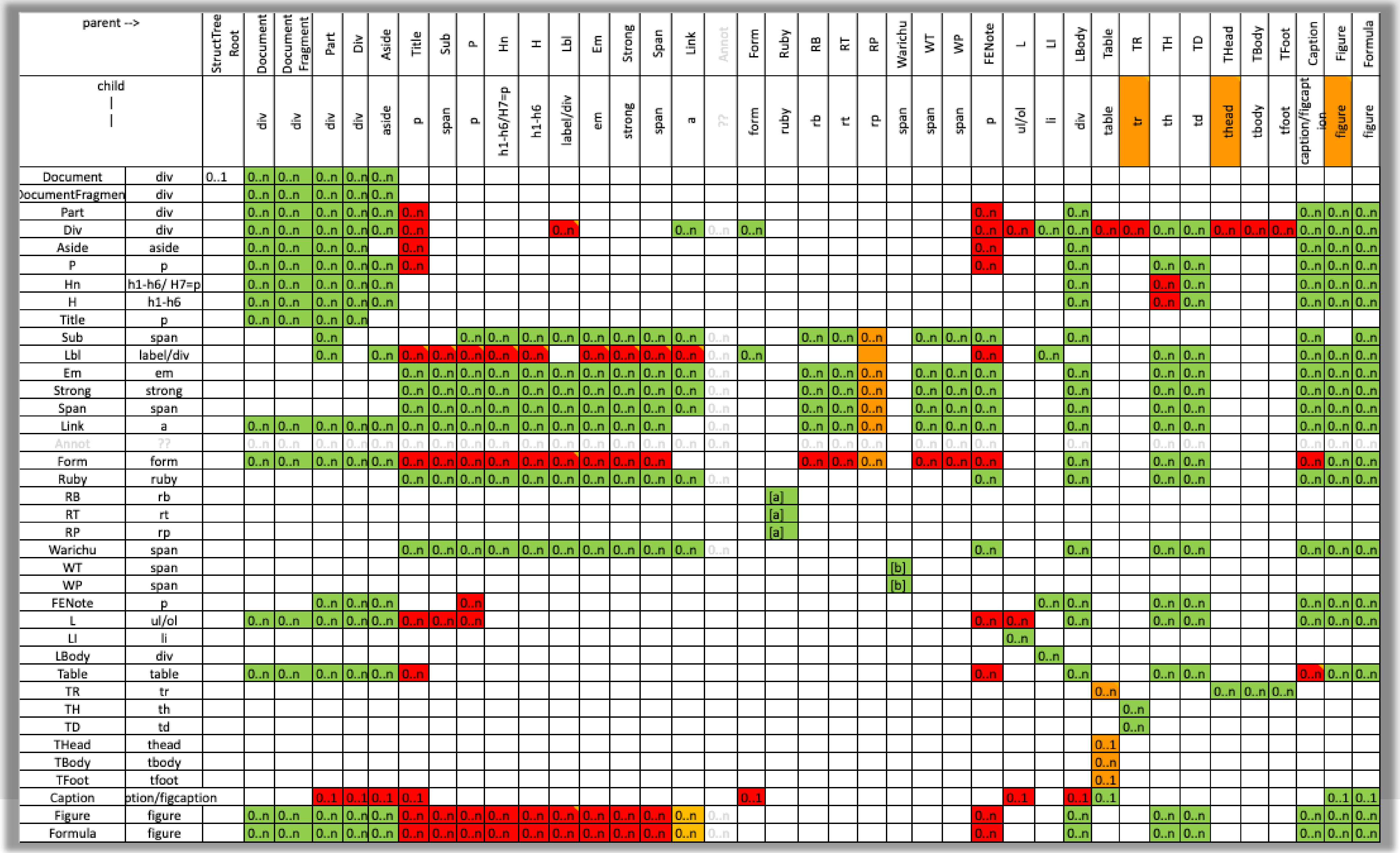
Submit form data to google spreadsheet.

Click here to check results

```
    this.getField("text_weight").value = weight*2.205;
}

}
all_fields.push({name:"group_gender", id:"28", annots:[]})
all_fields.push({name:"group_gender", id:"26", annots:[]})
all_fields.push({name:"listbox_ethnic", id:"27", annots:[]
all_fields.push({name:"text_age", id:"40", annots:[]});
function K40() {
AFNumber_Keystroke(0, 0, 0, 0, "", true);
}
function F40() {
AFNumber_Format(0, 0, 0, 0, "", true);
}
function B140() {

if(this.getField("text_age").value >= 18){
  this.getField("checkbox_adult").checkThisBox(0,true);
}
else{
  this.getField("checkbox_adult").checkThisBox(0,false);
}

}
all_fields.push({name:"checkbox_adult", id:"29", annots:[]
all_fields.push({name:"text_hip", id:"30", annots:[]});
function K30() {
AFNumber_Keystroke(1, 0, 0, 0, "", true);
}
function F30() {
AFNumber_Format(1, 0, 0, 0, "", true);
}
all_fields.push({name:"dropdown_hip", id:"31", annots:[]})
function V31() {
var hip = this.getField("text_hip").value;
if(hip > 0){
  if(event.value=="cm")
    this.getField("text_hip").value = hip*2.54;
  else
    this.getField("text_hip").value = hip/2.54;
}

}
all_fields.push({name:"text_bmi", id:"32", annots:[]});
function K32() {
AFNumber_Keystroke(2, 0, 0, 0, "", true);
}
function F32() {
AFNumber_Format(2, 0, 0, 0, "", true);
```

Line 1, Column 1

Valid html



parent -->		StructTree Root	Document	Document Fragment	Part	Div	Aside	Title	Sub	P	Hn	H	Lbl	Em	Strong	Span	Link	Annot	Form	Ruby	RB	RT	RP	Warichu	WT	WP	FENote	L	LI	LBody	Table	TR	TH	TD	THead	TBody	TFoot	Caption	Figure	Formula
child			div	div	div	div	aside	p	span	p	h1-h6/H7=p	h1-h6	label/div	em	strong	span	a	??	form	ruby	rb	rt	rp	span	span	span	p	ul/ol	li	div	table	tr	th	td	thead	tbody	tfoot	caption/figcaption	figure	figure
Document	div	0..1	0..n	0..n	0..n	0..n	0..n																																	
DocumentFragmen	div		0..n	0..n	0..n	0..n	0..n																																	
Part	div		0..n	0..n	0..n	0..n	0..n	0..n																			0..n			0..n								0..n	0..n	0..n
Div	div		0..n	0..n	0..n	0..n	0..n	0..n					0..n				0..n	0..n	0..n								0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
Aside	aside		0..n	0..n	0..n	0..n		0..n																			0..n			0..n								0..n	0..n	0..n
P	p		0..n	0..n	0..n	0..n	0..n	0..n																			0..n			0..n			0..n	0..n				0..n	0..n	0..n
Hn	h1-h6/ H7=p		0..n	0..n	0..n	0..n	0..n																							0..n			0..n	0..n				0..n	0..n	0..n
H	h1-h6		0..n	0..n	0..n	0..n	0..n																							0..n			0..n	0..n				0..n	0..n	0..n
Title	p		0..n	0..n	0..n	0..n																																		
Sub	span				0..n					0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n								0..n		0..n
Lbl	label/div				0..n		0..n	0..n	0..n	0..n	0..n	0..n		0..n	0..n	0..n	0..n	0..n	0..n								0..n		0..n				0..n	0..n				0..n	0..n	0..n
Em	em							0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Strong	strong							0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Span	span							0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Link	a		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n		0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Annot	??		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Form	form		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n					0..n	0..n	0..n		0..n	0..n	0..n			0..n			0..n	0..n				0..n	0..n	0..n
Ruby	ruby							0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n									0..n			0..n			0..n	0..n				0..n	0..n	0..n
RB	rb																			[a]																				
RT	rt																			[a]																				
RP	rp																			[a]																				
Warichu	span							0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n									0..n			0..n			0..n	0..n				0..n	0..n	0..n
WT	span																							[b]																
WP	span																							[b]																
FENote	p				0..n	0..n	0..n			0..n																			0..n	0..n			0..n	0..n				0..n	0..n	0..n
L	ul/ol		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n																	0..n	0..n		0..n			0..n	0..n				0..n	0..n	0..n
LI	li																											0..n												
LBody	div																												0..n											
Table	table		0..n	0..n	0..n	0..n	0..n	0..n																			0..n			0..n			0..n	0..n				0..n	0..n	0..n
TR	tr																														0..n				0..n	0..n	0..n			
TH	th																															0..n								
TD	td																															0..n								
THead	thead																														0..1									
TBody	tbody																														0..n									
TFoot	tfoot																														0..1									
Caption	ption/figcaption				0..1	0..1	0..1	0..1											0..1									0..1		0..1	0..1								0..1	0..1
Figure	figure		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n									0..n			0..n			0..n	0..n				0..n	0..n	0..n
Formula	figure		0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n									0..n			0..n			0..n	0..n				0..n	0..n	0..n

Roman Toda NORMEX

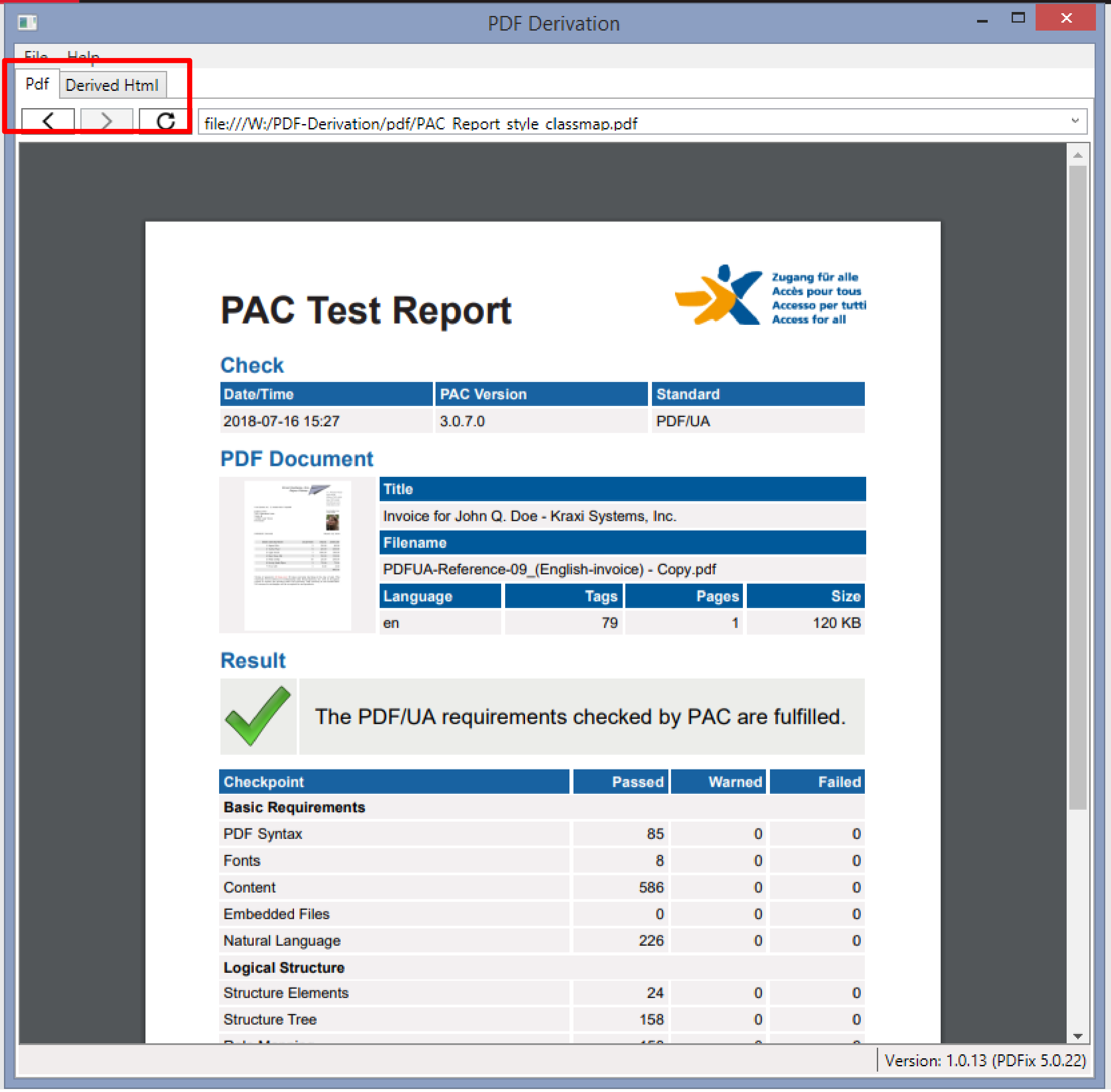
Summary

- World of single representation is over. Today we need to read out loud our pdfs, but what’s going to be tomorrow?
- It’s time to stop producing bad pdfs
- It’s time to give authors full control over pdf interpretation
- Author decides instead of tool decides
- Instead of saying there is only one way to interpret PDF we will be saying that there is **always deterministic** way of interpreting pdf

Resources

- https://github.com/Normex/PDF-Derivation
- Sample files
  - Styling
  - Associated files
  - Forms
  - Interactive
  - Fail cases
- Implementation (commandline, GUI)

Resources



file:///W:/PDF-Derivation/pdf/PAC Report style classmap.pdf

PAC Test Report

Check

Date/Time	PAC Version	Standard
2018-07-16 15:27	3.0.7.0	PDF/UA

PDF Document

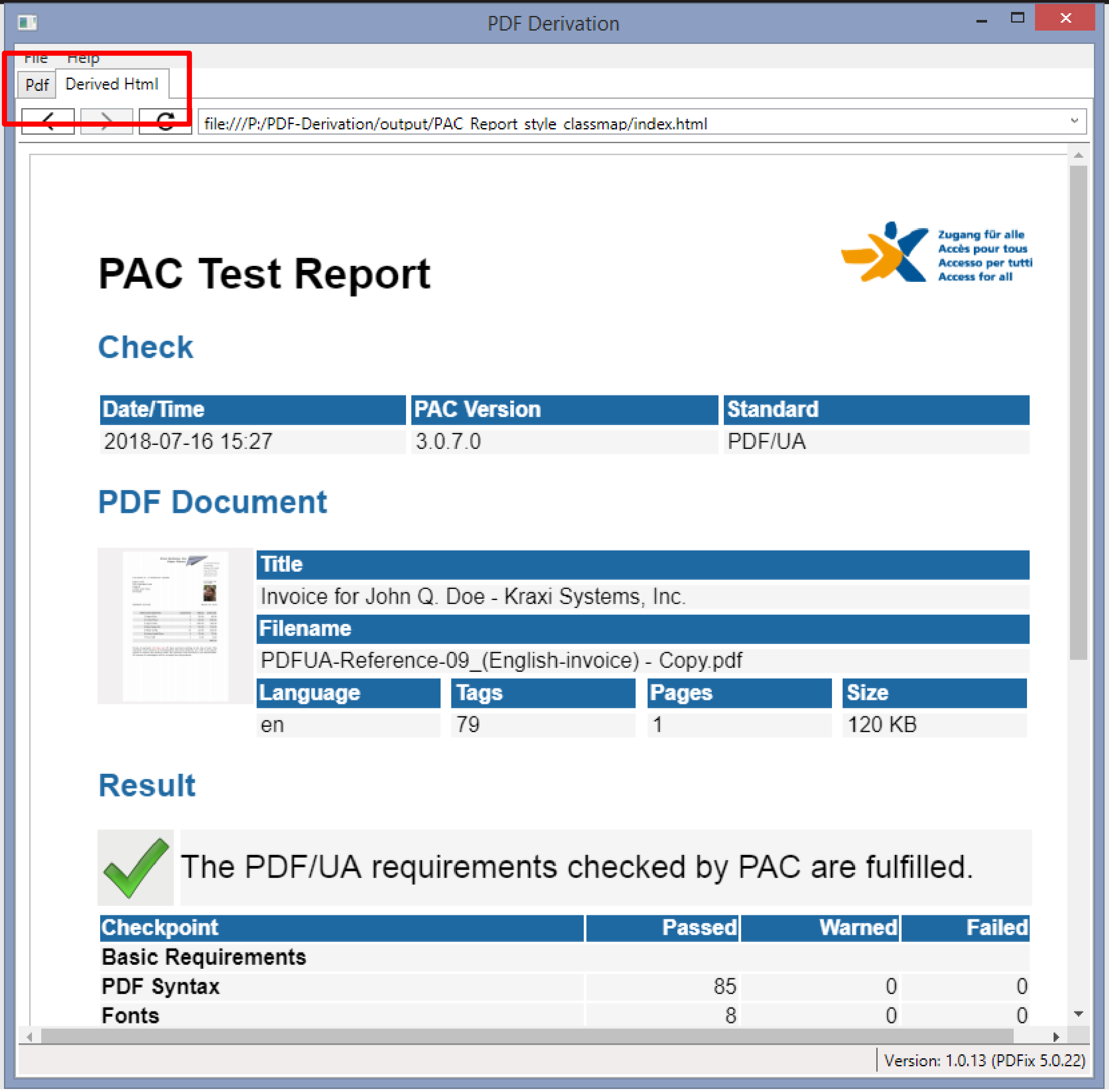
Title			
Invoice for John Q. Doe - Kraxi Systems, Inc.			
Filename			
PDFUA-Reference-09_(English-invoice) - Copy.pdf			
Language	**Tags**	**Pages**	**Size**
en	79	1	120 KB

Result

✓ The PDF/UA requirements checked by PAC are fulfilled.

Checkpoint	Passed	Warned	Failed
Basic Requirements			
PDF Syntax	85	0	0
Fonts	8	0	0
Content	586	0	0
Embedded Files	0	0	0
Natural Language	226	0	0
Logical Structure			
Structure Elements	24	0	0
Structure Tree	158	0	0

Version: 1.0.13 (PDFix 5.0.22)



file:///P:/PDF-Derivation/output/PAC Report style classmap/index.html

PAC Test Report

Check

Date/Time	PAC Version	Standard
2018-07-16 15:27	3.0.7.0	PDF/UA

PDF Document

Title			
Invoice for John Q. Doe - Kraxi Systems, Inc.			
Filename			
PDFUA-Reference-09_(English-invoice) - Copy.pdf			
Language	**Tags**	**Pages**	**Size**
en	79	1	120 KB

Result

✓ The PDF/UA requirements checked by PAC are fulfilled.

Checkpoint	Passed	Warned	Failed
Basic Requirements			
PDF Syntax	85	0	0
Fonts	8	0	0

Version: 1.0.13 (PDFix 5.0.22)

?

Thanks !

Roman Toda, Normex

https://github.com/Normex/PDF-Derivation