# Aufreinigung von OCR-erkannten Texten mit Python

Adrian Pachzelt

Universitätsbibliothek J. C. Senckenberg

# Das Ausgangsproblem

## Original Seite

### Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

*Our planet photographed from "Apollo - 17". You can notice the big ice polar cap in Antarctica.*

*Grand Canyon: the stream cut through the layers of soft sandstone and limestone in Arizona (USA) into a gigantic valley.*
*The maximum depth is 1.9 km.*

*Mississippi Delta: the river Mississippi collects a great amount of sediments along its long run and then drops them in the Gulf of Mexico. This slow river leaves these sediments in the mouth, thus creating areas of new land that didn't exist before.*

**Seasons**

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

**Structure**

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter substances that form the crust. The atmosphere is made of nitrogen (77%), oxygen (21%), and a mixture of water vapor and other gases.

**Magnetic bubble**

The rotation of Earth around its axis generates forceful electrical currents in the iron core of the planet and this creates the magnetic field. This field forms a giant "bubble" in the near-Earth space called the "magnetosphere" Magnetosphere protects Earth from the solar wind – a flow of charged particles emitted by the Sun. These particles are trapped by the magnetic field in two huge rings – Van Allen's belts. When spacecrafts travel through the Van Allen's belts, the electrical equipment of the former may suffer malfunction caused by these particles.

**Clashing Continents**

The Earth crust is made from parts called plates, which float on its surface driven by the flows in the liquid mantle. The continents lie on these plates, and so their location is subject to constant change. Some 200 million years ago, all the dry land on Earth was a single continent called Pangea by the scientists, which further split into the continents we know now. The lava rises by millimeters around the mountain ridges located on the ocean floor, and moves the continents apart. When the continents clash, as they do around the shores of the Pacific, the Earth's surface swells up and mountain ridges rise up; if the plates go down into the mantle, earthquakes and volcanic eruptions occur. This process, called plate tectonics, constantly changes the Earth's appearance.

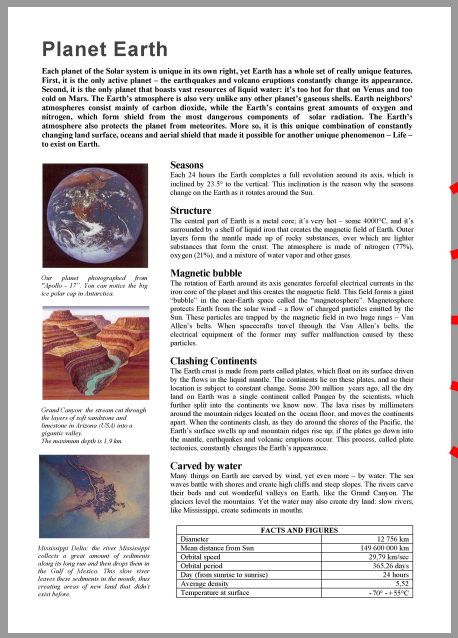
**Carved by water**

Many things on Earth are carved by wind, yet even more – by water. The sea waves battle with shores and create high cliffs and steep slopes. The rivers carve their beds and cut wonderful valleys on Earth, like the Grand Canyon. The glaciers level the mountains. Yet the water may also create dry land: slow rivers, like Mississippi, create sediments in mouths.

| FACTS AND FIGURES | |
|---|---|
| Diameter | 12 756 km |
| Mean distance from Sun | 149 600 000 km |
| Orbital speed | 29.79 km/sec |
| Orbital period | 365.26 days |
| Day (from sunrise to sunrise) | 24 hours |
| Average density | 5.52 |
| Temperature at surface | -70° - + 55°C |

# Das Ausgangsproblem

## Original Seite



**Planet Earth**

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

*Our planet photographed from "Apollo - 17". You can notice the big ice polar cap in Antarctica.*

*Grand Canyon: the stream cut through the layers of soft sandstone and limestone in Arizona (USA) into a gigantic valley.*
*The maximum depth is 1,9 km.*

*Mississippi Delta: the river Mississippi collects a great amount of sediments along its long run and then drops them in the Gulf of Mexico. This slow river leaves these sediments in the mouth, thus creating areas of new land that didn't exist before.*

**Seasons**
Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

**Structure**
The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter substances that form the crust. The atmosphere is made of nitrogen (77%), oxygen (21%), and a mixture of water vapor and other gases.

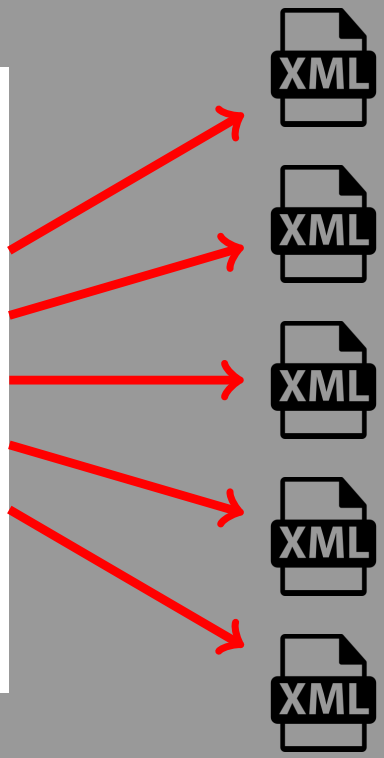
**Magnetic bubble**
The rotation of Earth around its axis generates forceful electrical currents in the iron core of the planet and this creates the magnetic field. This field forms a giant "bubble" in the near-Earth space called the "magnetosphere". Magnetosphere protects Earth from the solar wind – a flow of charged particles emitted by the Sun. These particles are trapped by the magnetic field in two huge rings – Van Allen's belts. When spacecrafts travel through the Van Allen's belts, the electrical equipment of the former may suffer malfunction caused by these particles.

**Clashing Continents**
The Earth crust is made from parts called plates, which float on its surface driven by the flows in the liquid mantle. The continents lie on these plates, and so their location is subject to constant change. Some 200 million years ago, all the dry land on Earth was a single continent called Pangea by the scientists, which further split into the continents we know now. The lava rises by millimeters around the mountain ridges located on the ocean floor, and moves the continents apart. When the continents clash, as they do around the shores of the Pacific, the Earth's surface swells up and mountain ridges rise up; if the plates go down into the mantle, earthquakes and volcanic eruptions occur. This process, called plate tectonics, constantly changes the Earth's appearance.

**Carved by water**
Many things on Earth are carved by wind, yet even more – by water. The sea waves battle with shores and create high cliffs and steep slopes. The rivers carve their beds and cut wonderful valleys on Earth, like the Grand Canyon. The glaciers level the mountains. Yet the water may also create dry land: slow rivers, like Mississippi, create sediments in mouths.

| FACTS AND FIGURES | |
|---|---|
| Diameter | 12 756 km |
| Mean distance from Sun | 149 600 000 km |
| Orbital speed | 29,79 km/sec |
| Orbital period | 365,26 days |
| Day (from sunrise to sunrise) | 24 hours |
| Average density | 5,52 |
| Temperature at surface | - 70° - + 55°C |

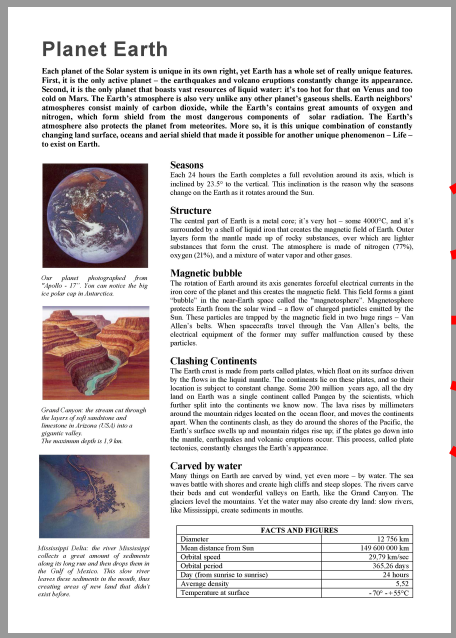
## ABBYY Output




```
<line>Das ist <a id='test'>
ein <b>fettes</b>
Beispiel</a>.</line>
```

# Das Ausgangsproblem

Original Seite ABBYY Output Klartext



## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

*Our planet photographed from "Apollo - 17". You can notice the big ice polar cap in Antarctica.*

*Grand Canyon: the stream cut through the layers of soft sandstone and limestone in Arizona (USA) into a gigantic valley.*
*The maximum depth is 1,9 km.*

*Mississippi Delta: the river Mississippi collects a great amount of sediments along its long run and then drops them in the Gulf of Mexico. This slow river leaves these sediments in the mouth, thus creating areas of new land that didn't exist before.*

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core: it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter substances that form the crust. The atmosphere is made of nitrogen (77%), oxygen (21%), and a mixture of water vapor and other gases.

### Magnetic bubble

The rotation of Earth around its axis generates forceful electrical currents in the iron core of the planet and this creates the magnetic field. This field forms a giant "bubble" in the near-Earth space called the "magnetosphere". Magnetosphere protects Earth from the solar wind – a flow of charged particles emitted by the Sun. These particles are trapped by the magnetic field in two huge rings – Van Allen's belts. When spacecrafts travel through the Van Allen's belts, the electrical equipment of the former may suffer malfunction caused by these particles.

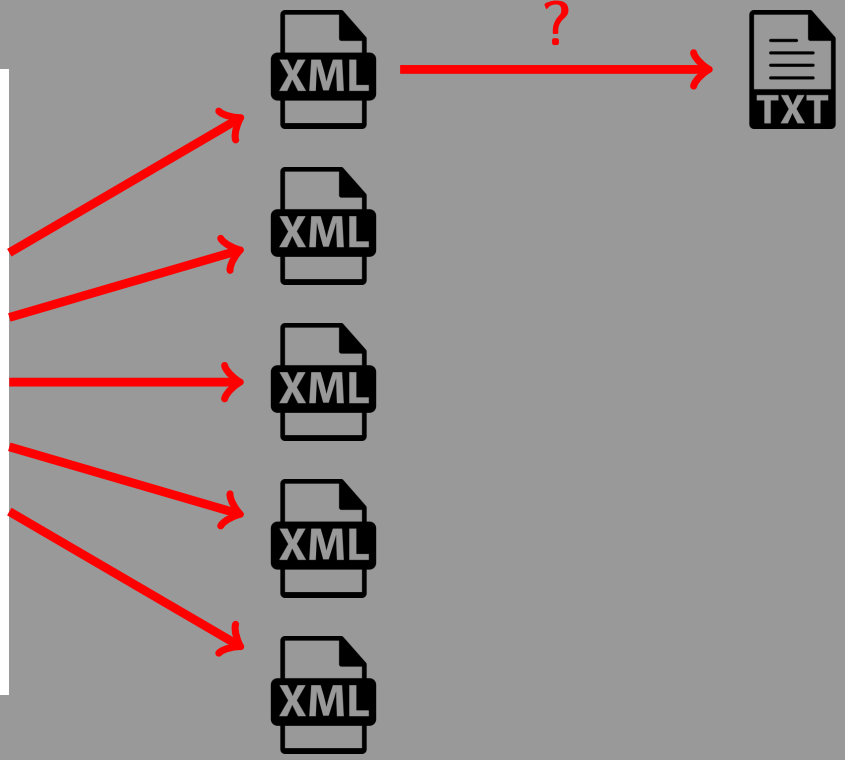
### Clashing Continents

The Earth crust is made from parts called plates, which float on its surface driven by the flows in the liquid mantle. The continents lie on these plates, and so their location is subject to constant change. Some 200 million years ago, all the dry land on Earth was a single continent called Pangea by the scientists, which further split into the continents we know now. The lava rises by millimeters around the mountain ridges located on the ocean floor, and moves the continents apart. When the continents clash, as they do around the shores of the Pacific, the Earth's surface swells up and mountain ridges rise up; if the plates go down into the mantle, earthquakes and volcanic eruptions occur. This process, called plate tectonics, constantly changes the Earth's appearance.

### Carved by water

Many things on Earth are carved by wind, yet even more – by water. The sea waves battle with shores and create high cliffs and steep slopes. The rivers carve their beds and cut wonderful valleys on Earth, like the Grand Canyon. The glaciers level the mountains. Yet the water may also create dry land: slow rivers, like Mississippi, create sediments in mouths.

| FACTS AND FIGURES | |
|---|---|
| Diameter | 12 756 km |
| Mean distance from Sun | 149 600 000 km |
| Orbital speed | 29,79 km/sec |
| Orbital period | 365,26 days |
| Day (from sunrise to sunrise) | 24 hours |
| Average density | 5,52 |
| Temperature at surface | - 70° - + 55°C |



# Und so wird's gemacht!

```
from bs4 import BeautifulSoup as Soup
with open("/user/abbyy.xml") as ifile:
        xml_data = ifile.read()
xml_data = Soup(xml_text, "lxml")
print(xml_data.text)
```

# Ja nee is klar!

# Original Seite

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

*Our planet photographed from "Apollo - 17". You can notice the big ice polar cap in Antarctica.*

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter substances that form the crust. The atmosphere is made of nitrogen (77%), oxygen (21%), and a mixture of water vapor and other gases.

### Magnetic bubble

The rotation of Earth around its axis generates forceful electrical currents in the iron core of the planet and this creates the magnetic field. This field forms a giant "bubble" in the near-Earth space called the "magnetosphere". Magnetosphere protects Earth from the solar wind – a flow of charged particles emitted by the Sun. These particles are trapped by the magnetic field in two huge rings – Van Allen's belts. When spacecrafts travel through the Van Allen's belts, the electrical equipment of the former may suffer malfunction caused by these

# ABBYY XML

<page width="2479" height="3508" resolution="300"> <block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183" t="187" r="2269" b="789"><region><rect l="183" t="187" r="2269" b="789"/></region> <text id="678B5513-5619-4752-92A0-48F4CED1B777"> <par align="Justified" leftIndent="100"> <line baseline="275" l="196" t="192" r="888" b="276"><formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20"> <charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams> <charParams l="276" t="192" r="292" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="255" wordPenalty="0" meanStrokeWidth="167">I</charParams> <charParams l="306" t="212" r="362" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="167">a</charParams> <charParams l="378" t="212" r="434" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="167">n</charParams> <charParams l="448" t="212" r="504" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="40" wordPenalty="0" meanStrokeWidth="167">e</charParams> <charParams l="512" t="194" r="550" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="28" wordPenalty="0" meanStrokeWidth="167">t</charParams> <charParams l="551" t="192" r="595" b="276"> </charParams> <charParams l="596" t="192" r="658" b="274" wordStart="1" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="171">E</charParams> <charParams l="672" t="212" r="726" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="171">a</charParams> <charParams l="744" t="212" r="780" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="4" wordPenalty="0" meanStrokeWidth="171">r</charParams>

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

Breite

Höhe

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

Indent ⟷

## Planet Earth

Justified = "Blocksatz"

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

## Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

### Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

### Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```

# Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**

## Seasons

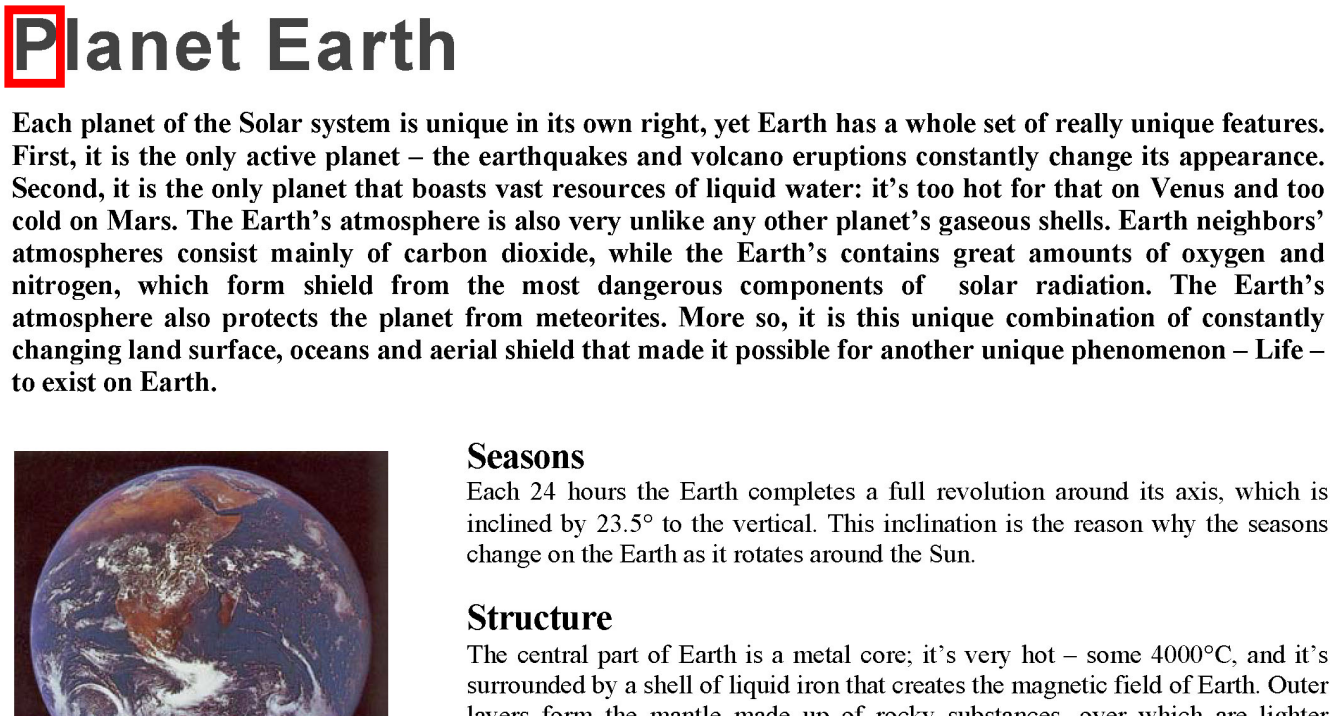
Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

## Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer layers form the mantle made up of rocky substances, over which are lighter

# ABBYY XML

```
<page width="2479" height="3508" resolution="300">
<block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183"
t="187" r="2269" b="789">
<region><rect l="183" t="187" r="2269" b="789"/></region>
<text id="678B5513-5619-4752-92A0-48F4CED1B777">
<par align="Justified" leftIndent="100">
<line baseline="275" l="196" t="192" r="888" b="276">
<formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20">
<charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1"
wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100"
serifProbability="4" wordPenalty="0" meanStrokeWidth="167">P</charParams>
```



# Planet Earth

**Each planet of the Solar system is unique in its own right, yet Earth has a whole set of really unique features. First, it is the only active planet – the earthquakes and volcano eruptions constantly change its appearance. Second, it is the only planet that boasts vast resources of liquid water: it's too hot for that on Venus and too cold on Mars. The Earth's atmosphere is also very unlike any other planet's gaseous shells. Earth neighbors' atmospheres consist mainly of carbon dioxide, while the Earth's contains great amounts of oxygen and nitrogen, which form shield from the most dangerous components of solar radiation. The Earth's atmosphere also protects the planet from meteorites. More so, it is this unique combination of constantly changing land surface, oceans and aerial shield that made it possible for another unique phenomenon – Life – to exist on Earth.**



## Seasons

Each 24 hours the Earth completes a full revolution around its axis, which is inclined by 23.5° to the vertical. This inclination is the reason why the seasons change on the Earth as it rotates around the Sun.

## Structure

The central part of Earth is a metal core; it's very hot – some 4000°C, and it's surrounded by a shell of liquid iron that creates the magnetic field of Earth. Outer

# ABBYY XML

<page width="2479" height="3508" resolution="300"> <block blockType="Text" pageElemId="678B5513-5619-4752-92A0-48F4CED1B777" l="183" t="187" r="2269" b="789"><region><rect l="183" t="187" r="2269" b="789"/></region> <text id="678B5513-5619-4752-92A0-48F4CED1B777"> <par align="Justified" leftIndent="100"> <line baseline="275" l="196" t="192" r="888" b="276"><formatting lang="EnglishUnitedStates" ff="Arial" fs="27.5" bold="1" spacing="20"> <charParams l="196" t="192" r="260" b="274" wordStart="1" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="4" wordPenalty="0" meanStrokeWidth="167">**P**</charParams> <charParams l="276" t="192" r="292" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="255" wordPenalty="0" meanStrokeWidth="167">**l**</charParams> <charParams l="306" t="212" r="362" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="167">**a**</charParams> <charParams l="378" t="212" r="434" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="167">**n**</charParams> <charParams l="448" t="212" r="504" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="40" wordPenalty="0" meanStrokeWidth="167">**e**</charParams> <charParams l="512" t="194" r="550" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="28" wordPenalty="0" meanStrokeWidth="167">**t**</charParams> <charParams l="551" t="192" r="595" b="276">_</charParams> <charParams l="596" t="192" r="658" b="274" wordStart="1" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="171">**E**</charParams> <charParams l="672" t="212" r="726" b="276" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="0" wordPenalty="0" meanStrokeWidth="171">**a**</charParams> <charParams l="744" t="212" r="780" b="274" wordStart="0" wordFromDictionary="1" wordNormal="1" wordNumeric="0" wordIdentifier="0" charConfidence="100" serifProbability="4" wordPenalty="0" meanStrokeWidth="171">**r**</charParams>

# Python3 to the rescue!

# Warum eigentlich Python?

- Alternativen: Java, C/C++, JavaScript
  - Können alles und mehr...
  - Suuuuper schnell!
  - **Einfach sich in den Fuß zu schießen!**
- Schnell und einfach zu lernen
  - Semantik ist an natürliche Sprache angelehnt
  - Ästhetische Syntax
  - Keine Typdeklaration bei Variablen

```
time = 1
time = "DAS! IST! PYTHON!"
```

- Schnelle Installation und `import` von Modulen

```
import json
from json import dump
from json import dump as Dumper
```

# Warum eigentlich Python?

- Alternativen: Java, C/C++, JavaScript
  - Können alles und mehr...
  - Suuuuper schnell!
  - **Einfach sich in den Fuß zu schießen!**
- Schnell und einfach zu lernen
  - Semantik ist an natürliche Sprache angelehnt
  - Ästhetische Syntax
  - Keine Typdeklaration bei Variablen

```
time = 1
time = "DAS! IST! PYTHON!"
```

- Schnelle Installation und `import` von Modulen

```
import json
from json import dump
from json import dump as Dumper
```

# Warum eigentlich Python?

- Alternativen: Java, C/C++, JavaScript
  - Können alles und mehr...
  - Suuuuper schnell!
  - **Einfach sich in den Fuß zu schießen!**
- Schnell und einfach zu lernen
  - Semantik ist an natürliche Sprache angelehnt
  - Ästhetische Syntax
  - Keine Typdeklaration bei Variablen

```
time = 1
time = "DAS! IST! PYTHON!"
```

- Schnelle Installation und `import` von Modulen

```
import json
from json import dump
from json import dump as Dumper
```

# Python 101

Für die Behandlung von ABBYY-XML brauchen wir:

- **For-Schleifen**: Wiederholen von Befehlen
- **Funktionen** aufrufen: Das coole Zeug!

# Die For-Schleife

```
lst = ["Das ", "ist ", "ein ", "Text."]
```

```
for t in lst:
    print(t)
```

---

**Ausgabe:** "Das ist ein Text."

## For-Schleifen Syntax

**for** Variablenname innerhalb der Schleife **in** Container (z.B. Liste):
    *# Code innerhalb der Schleife*
    *# Noch mehr Code innerhalb der Schleife*
*# Code außerhalb der Schleife*

# Die For-Schleife

```
lst = ["Das ", "ist ", "ein ", "Text."]
```

```
for t in lst:
    print(t)
```

---

**Ausgabe:** "Das ist ein Text."

## For-Schleifen Syntax

**for** Variablenname innerhalb der Schleife **in** Container (z.B. Liste):
    *# Code innerhalb der Schleife*
    *# Noch mehr Code innerhalb der Schleife*
*# Code außerhalb der Schleife*

# Die For-Schleife

```
lst = ["Das ", "ist ", "ein ", "Text."]
```

```
for t in lst:
    print(t)
```

---

**Ausgabe:** "Das ist ein Text."

## For-Schleifen Syntax

**for** Variablenname innerhalb der Schleife **in** Container (z.B. Liste):
    *# Code innerhalb der Schleife*
    *# Noch mehr Code innerhalb der Schleife*
*# Code außerhalb der Schleife*

# Die For-Schleife

```
lst = ["Das ", "ist ", "ein ", "Text."]
```

```
for t in lst:
    print(t)
```

---

**Ausgabe:** "Das ist ein Text."

## For-Schleifen Syntax

**for** Variablenname innerhalb der Schleife **in** Container (z.B. Liste):
    *# Code innerhalb der Schleife*
    *# Noch mehr Code innerhalb der Schleife*
*# Code außerhalb der Schleife*

# Funktionen aufrufen

```
text = "das ist ein text."
```

# Funktionen aufrufen

```
text = "das ist ein text."

text.capitalize() # "Das ist ein text."
```

# Funktionen aufrufen

```
text = "das ist ein text."

text.capitalize() # "Das ist ein text."

text.upper() # "DAS IST EIN TEXT."
```

# Funktionen aufrufen

```
text = "das ist ein text."

text.capitalize() # "Das ist ein text."

text.upper() # "DAS IST EIN TEXT."

text.find("i") # 4
text.find("l") # -1
```

# Funktionen aufrufen

```
text = "das ist ein text."

text.capitalize() # "Das ist ein text."

text.upper() # "DAS IST EIN TEXT."

text.find("i") # 4
text.find("l") # -1

with open("text.txt", "w") as output:
    output.write(text)
```

**Funktionsaufruf Syntax**

Variablenname.Funktionsname(Parameter1, Parameter2, ...)

# Funktionen aufrufen

```
text = "das ist ein text."

text.capitalize() # "Das ist ein text."

text.upper() # "DAS IST EIN TEXT."

text.find("i") # 4
text.find("l") # -1

with open("text.txt", "w") as output:
    output.write(text)
```

## Funktionsaufruf Syntax

Variablenname.Funktionsname(Parameter1, Parameter2, ...)

# Livedemo-Time!

# Bildnachweis

**XML-Symbol**, **TXT-Symbol**: Icon made by *flaticon* from `www.flaticon.com`

**Atze Schröder**: `https://en.wikipedia.org/w/index.php?title=Atze_Schr%C3%B6der&oldid=879008022` (Abgerufen: 18. Januar 2019, 12:23 UTC)