

Wikipedia in Edurep



Wim Muskee

kennisnet :: edurep :: geekness

<http://flic.kr/p/7AXNfY>



Edurep?

maakt leermiddelen vindbaar



aanbieders leermiddelen
webpagina's, online documenten, boeken,
leskisten

metadata



WEBDESIGN RULES!

DE ONBETWISTE REGELS VAN HET WEB

(EN HUN TEGENARGUMENTEN)

Webdesign Rules is het handboek bij uitstek voor iedereen die betrokken is bij de ontwikkeling van een website. Met name voor communicatiebureaus en marketeers is Webdesign Rules een must.

WEBDESIGN RULES IS EEN BOEK DAT

- de benodigde voorkennis geeft, zonder daarbij in overbodige technische details te treden;
- rekening houdt met verschillende perspectieven;
- onnodige vertragingen en meningsbijgaven van budgetten helpt voorkomen.

Webdesign Rules stelt opdrachtgevers in staat om zonder technische achtergrond mee te denken bij ontwerpprocessen.

OVER DE AUTEUR

Stefan Bos werkt als front-end developer bij full-service internetbureau Mangrove. Hij is ook en schreef de online journalist onder andere voor Samen Uitgaan.

Categorie: Webdesign en Webwet
www.webdesignrules.nl
www.mangrove.nl



titel

beschrijving

auteur

uitgever

identificer

formaat



vakgebied



leerniveau



leerdoel



studielast

metadata verzamelen





indexeren, zoekbaar maken

API, voorbereiden voor vinden





VERMONT BREAD COMPANY
100% WHOLE WHEAT
HOTDOG BUNS

VERMONT BREAD COMPANY
100% WHOLE WHEAT
HOTDOG BUNS

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

VERMONT BREAD COMPANY
HOTDOG BUNS

VERMONT BREAD COMPANY
POTATO
HOTDOG BUNS

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
8 Sandwich Rolls
Famous Pastry Shoppe Inc.

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Martin's
Famous Dutch
POTATO
Long Roll

Organic Cottage Tin
£1.00 300g
UK 4

Organic Ciabatta
£1.50 300g
UK 4
100% Ciabatta Flour, 5% Semolina, Extra Virgin Olive Oil, Water, Yeast, Salt

Olive Focaccia
£2.50

Olive & Herb Focaccia
£2.50







ROLLS

The best way to lose weight and stay healthy is to decrease calorie intake while increasing calorie expenditure.

WHITE

Je bent hier: [Wikiwijs Home](#) » [vo](#) » Uitgebreid zoeken

Je hebt gezocht op: [alles wissen](#)

Kosten:
Nee [wis](#)

Overig:
Alleen digitaal leermateriaal [wis](#)

[bekijk 30307 resultaten](#)

Trefwoord

[voeg toe](#)

- [algemeen](#) [vakinhoud](#) [tips](#)

Selecteer hieronder de filters die je wilt toepassen.

Vakgebied ?

- Algemeen vormende vakken
 - Aardrijkskunde
 - Algemene natuurwetenschappen
 - Arabisch
 - Beeldende vorming
 - Biologie
 - Chinees
 - Culturele en kunstzinnige vorming
 - Dans
 - Drama
 - Duits

Kosten ?

- Ja
- Nee

Leerniveau ?

- HAVO
 - HAVO, 1
 - HAVO, 2
 - HAVO, 3
 - HAVO, 4
 - HAVO, 5
- Praktijkonderwijs
- VMBO
- VWO

Bron ?

- Educatheek.nl
- Nieuws in de klas

Kerdoel ?

- Bewegen en sport
- Engels

1 Elke nacht haalt Edurep alle gegevens (metadata) van aangesloten collecties op.

2 Gebruikers bezoeken een aangesloten website of leeromgeving en zoeken naar leermateriaal.



4 In het resultaat staat de link naar het gevonden materiaal. Door op de link te klikken haal je het materiaal rechtstreeks bij de aanbieder op.

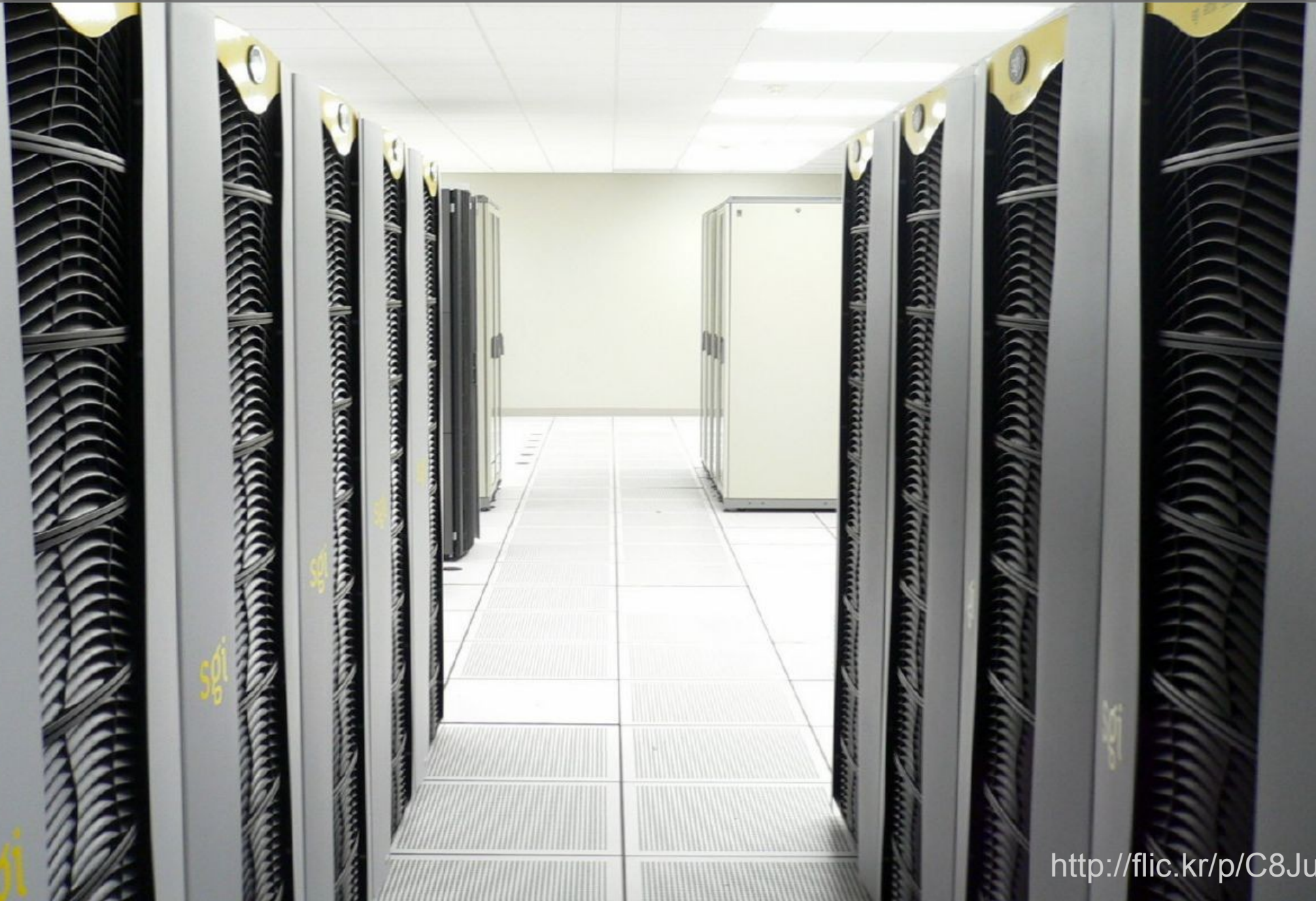
3 De website of leeromgeving geeft de zoekvraag door aan Edurep, die vervolgens de zoekresultaten teruggeeft.

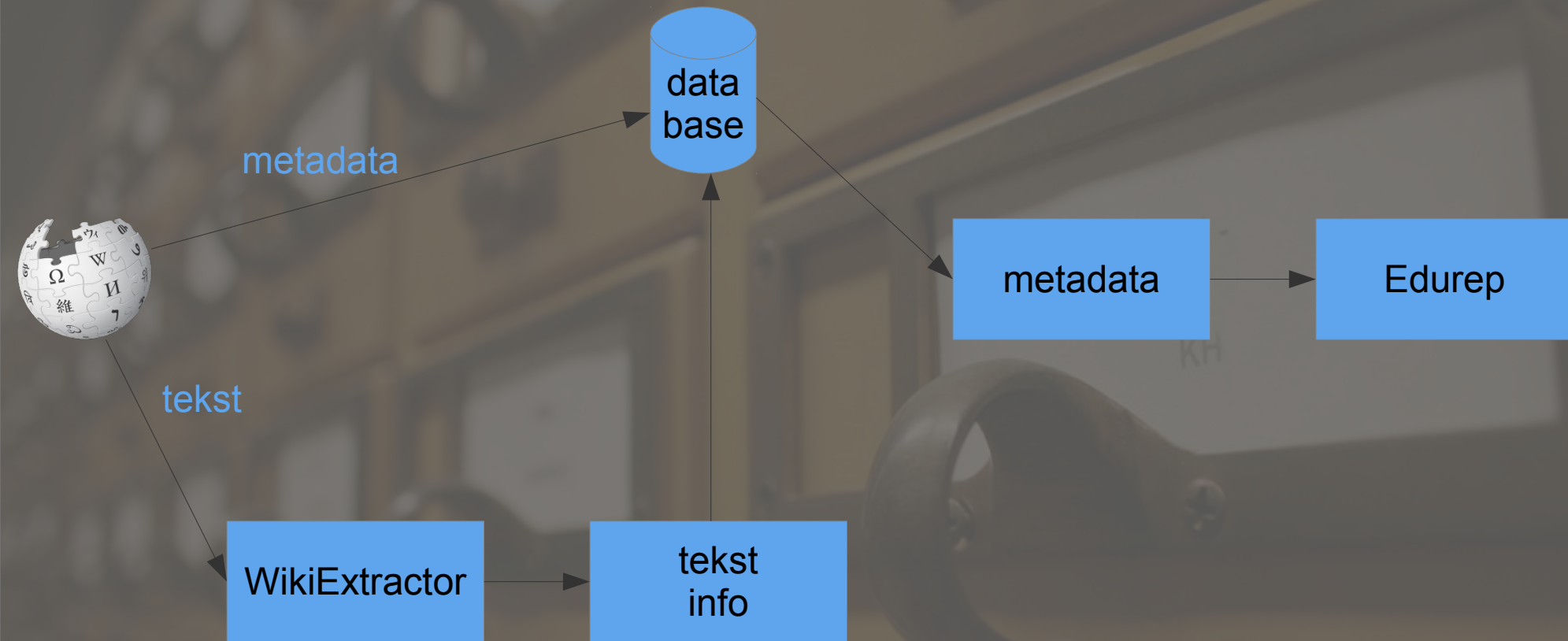


met de hand



automatiseren





Introduction

The project uses the *Italian Wikipedia* as source of documents for several purposes: as training data and as source of data to be annotated.

The Wikipedia maintainers provide, each month, an XML *dump* of all documents in the database: it consists of a single XML file containing the whole encyclopedia, that can be used for various kinds of analysis, such as statistics, service lists, etc.

Wikipedia dumps are available from [Wikipedia database download](#) .

The Wikipedia extractor tool generates plain text from a Wikipedia database dump, discarding any other information or annotation present in Wikipedia pages, such as images, tables, references and lists.

Each document in the dump of the encyclopedia is representend as a single XML element, encoded as illustrated in the following example from the document titled *Armonium*:

```
<page>
<title>Armonium</title>
<id>2</id>
<timestamp>2008-06-22T21:48:55Z</timestamp>
<username>Nemo bis</username>
<comment>italiano</comment>
<text xml:space="preserve">thumb|right|300 px

L'armonium' (in francese, harmonium) è uno
strumento musicale azionato con una tastiera, detta
manuale. Sono stati costruiti anche alcuni armonium con due manuali.

==Armonium occidentale==
Come l'organo, l'armonium è utilizzato tipicamente in
chiesa, per l'esecuzione di musica sacra, ed è
fornito di pochi registri, quando addirittura in certi casi non ne possiede
nemmeno uno: il suo timbro è molto meno ricco di quello
organistico e così pure la sua estensione.

...

==Armonium indiano==
Template:S sezione

== Voci correlate ==
*Musica
*Generi musicali</text>
```

```
</page></nowiki>
```

For this document the Wikipedia extractor produces the following plain text:

```
<doc id="2" url="http://it.wikipedia.org/wiki/Armonium">
Armonium.
L'armonium (in francese, "harmonium") è uno strumento musicale azionato con
una tastiera, detta manuale. Sono stati costruiti anche alcuni armonium con
due manuali.

Armonium occidentale.
Come l'organo, l'armonium è utilizzato tipicamente in chiesa, per l'esecuzione
di musica sacra, ed è fornito di pochi registri, quando addirittura in certi
casi non ne possiede nemmeno uno: il suo timbro è molto meno ricco di quello
organistico e così pure la sua estensione.
```

Contents [hide]

- 1 Introduction
- 2 Description
- 3 Example of Use
- 4 Downloads
- 5 Related Work

WEBDESIGN RULES!

DE ONBETWISTE REGELS VAN HET WEB

(EN HUN TEGENARGUMENTEN)

Webdesign Rules is het handboek bij uitstek voor iedereen die betrokken is bij de ontwikkeling van een website. Met name voor communicatiebureaus en marketeers is Webdesign Rules een must.

WEBDESIGN RULES IS EEN BOEK DAT

- de benodigde voorkennis geeft, zonder daarbij in overbodige technische details te treden;
- rekening houdt met verschillende perspectieven;
- onnodige vertragingen en meningsrijding van budgetten helpt voorkomen.

Webdesign Rules stelt opdrachtgevers in staat om zonder technische achtergrond mee te denken bij ontwerpprocessen.

OVER DE AUTEUR

Stefan Bos werkt als front-end developer bij full-service internetbureau Mangrove. Hij is ook en schreef de online publicatie onder andere voor Samen Uitgaan.

Categorie: Webdesign en Wereld
www.webdesignrules.nl
www.mangrove.nl



titel

beschrijving

trefwoord

identifier


```
wim@coruscant:/opt/Wikipedia/extract-old/AA * bzoat wiki_01.bz2 | head -n 10  
<doc id="81" url="http://nl.wikipedia.org/wiki/Atletiek" title="Atletiek">Atletiek
```

Atletiek is een oude sport waar de sporters (atleten) individueel of in groepen (estafette) moeten presteren. Atletiek wordt zowel op de weg als op een atletiekbaan beoefend. Atletiek wordt wel 'de moeder der sporten' genoemd omdat het de menselijke basisbewegingen (lopen, springen, werpen) omvat.

Atletiekbanen zijn meestal ovaal gevormd en 400 meter lang. Op het middenterrein worden de werp- en springonderdelen beoefend. Er zijn banen van gras, kunststof en gravel.

Dit artikel gaat over de atletiek als wedstrijdsport, maar atletiek in brede zin omvat ook recreatieve vormen. De Koninklijke Nederlandse Atletiek Unie (kortweg Atletiekunie) in Nederland doet ook veel voor de loopsport, sportief wandelen en Nordic walking. Voor de loopsporters zijn er wel wedstrijden maar ook de zogeheten trimlopen, 'wedstrijden' waarbij geen uitslag wordt opgemaakt. Slechts een klein deel van degenen die hardlopen als eerste sport noemen, is daadwerkelijk lid van de Atletiekunie.

Geschiedenis.

Atletiek is een sport die oorspronkelijk nauw verbonden was met de Griekse klassieke Olympische Spelen, die vanaf 776 voor Christus om de vier jaar gehouden werden als onderdeel van een feest ter ere van de god Zeus. Het vijfdaagse programma bestond uit sportieve krachtmetingen en wedstrijden in de schone kunsten. De beste atleten kwamen tegen elkaar uit in een vijfkamp, die bestond uit de onderdelen: hardlopen, verspringen, worstelen, speerwerpen en discuswerpen. De diverse onderdelen hadden veel minder regels dan nu.

Atletiek is samen met de zwemsport de oudste sport ter wereld. De Kretenzers (bewoners van Kreta) deden als eersten aan atletiek rond 1500 jaar v. Chr. De moderne atletiek ontstond in Engeland aan het einde van de 17e eeuw. De eerste professionele wedstrijden vonden daar plaats in de vroege 19e eeuw. Atletiek was een onderdeel op de eerste moderne Olympische Spelen in 1896. Nu is de atletiek nog steeds de belangrijkste sport in de Olympische Spelen, maar de band tussen atletiek en de Olympische Spelen is wel losser geworden. Dit kun je merken aan het feit dat naast de Olympische Spelen nog een heleboel andere atletiekevenementen worden gehouden, zoals de jaarlijkse Europacupwedstrijden en de Wereldkampioenschappen atletiek, die voor het eerst plaatsvonden in 1983.

Leeftijdsindeling.

Bij grote wedstrijden wordt niet naar leeftijd gekeken, het gaat puur om de prestatie. Slechts een enkeling is zo getalenteerd dat hij of zij al op jonge leeftijd om de prijzen mee kan doen en daarom zijn er ook wedstrijden en kampioenschappen naar leeftijd, waar meer met gelijken gestreden kan worden. Internationaal zijn atleten junior/jeugd tot en met het jaar van de negentiende verjaardag; de overgang naar de senioren vindt plaats op 1 januari van het jaar waarin men twintig wordt. De onderverdeling van de junioren is niet internationaal vastgelegd en verschilt van land tot land. In Nederland en Vlaanderen is de onderverdeling als volgt, waarbij niet naar de negentiende verjaardag wordt gekeken maar naar het jaar waarin men negentien wordt, enzovoort. De doorschuiving bij de categorieën vindt in Nederland plaats op 1 november. Bij de masters is dat op de verjaardag zelf.

```
wim@coruscant /opt/Wikipedia/extract-old/AA * []
```



```
wim@coruscant /opt/Wikipedia/extract-old/AA $ bzcatt wiki_01.bz2 | head -n 10 | grep -iw "sport\latletiek\lafstanden\latleten\wedstrijden\kampioenschappen\lolympische"
```

```
<doc id="81" url="http://nl.wikipedia.org/wiki/Atletiek" title="Atletiek">Atletiek
```

Atletiek is een oude **sport** waar de sporters (**atleten**) individueel of in groepen (estafette) moeten presteren. **Atletiek** wordt zowel op de weg als op een atletiekbaan beoefend. **Atletiek** wordt wel 'de moeder der sporten' genoemd omdat het de menselijke basisbewegingen (lopen, springen, werpen) omvat.

Dit artikel gaat over de **atletiek** als wedstrijdsport, maar **atletiek** in brede zin omvat ook recreatieve vormen. De Koninklijke Nederlandse **Atletiek** Unie (kortweg Atletiekunie) in Nederland doet ook veel voor de loopsport, sportief wandelen en Nordic walking. Voor de loopsporters zijn er wel **wedstrijden** maar ook de zogeheten trimlopen, '**wedstrijden**' waarbij geen uitslag wordt opgemaakt. Slechts een klein deel van degenen die hardlopen als eerste **sport** noemen, is daadwerkelijk lid van de Atletiekunie.

Atletiek is een **sport** die oorspronkelijk nauw verbonden was met de Griekse klassieke **Olympische** Spelen, die vanaf 776 voor Christus om de vier jaar gehouden werden als onderdeel van een feest ter ere van de god Zeus. Het vijfdaagse programma bestond uit sportieve krachtmetingen en **wedstrijden** in de schone kunsten. De beste **atleten** kwamen tegen elkaar uit in een vijfkamp, die bestond uit de onderdelen: hardlopen, verspringen, worstelen, speerwerpen en discuswerpen. De diverse onderdelen hadden veel minder regels dan nu.

Atletiek is samen met de zwemsport de oudste **sport** ter wereld. De Kretenzers (bewoners van Kreta) deden als eersten aan **atletiek** rond 1500 jaar v. Chr. De moderne **atletiek** ontstond in Engeland aan het einde van de 17e eeuw. De eerste professionele **wedstrijden** vonden daar plaats in de vroege 19e eeuw. **Atletiek** was een onderdeel op de eerste moderne **Olympische** Spelen in 1896. Nu is de **atletiek** nog steeds de belangrijkste **sport** in de **Olympische** Spelen, maar de band tussen **atletiek** en de **Olympische** Spelen is wel losser geworden. Dit kun je merken aan het feit dat naast de **Olympische** Spelen nog een heleboel andere atletiekevenementen worden gehouden, zoals de jaarlijkse Europacupwedstrijden en de Wereldkampioenschappen **atletiek**, die voor het eerst plaatsvonden in 1983.

Bij grote **wedstrijden** wordt niet naar leeftijd gekeken, het gaat puur om de prestatie. Slechts een enkeling is zo getalenteerd dat hij of zij al op jonge leeftijd om de prijzen mee kan doen en daarom zijn er ook **wedstrijden** en kampioenschappen naar leeftijd, waar meer met gelijken gestreden kan worden. Internationaal zijn **atleten** junior/jeugd tot en met het jaar van de negentiende verjaardag; de overgang naar de senioren vindt plaats op 1 januari van het jaar waarin men twintig wordt. De onderverdeling van de junioren is niet internationaal vastgelegd en verschilt van land tot land. In Nederland en Vlaanderen is de onderverdeling als volgt, waarbij niet naar de negentiende verjaardag wordt gekeken maar naar het jaar waarin men negentien wordt, enzovoort. De doorschuiving bij de categorieën vindt in Nederland plaats op 1 november. Bij de masters is dat op de verjaardag zelf.

```
wim@coruscant /opt/Wikipedia/extract-old/AA $
```




Serie elementen | **Bron:** [Wikipedia](#)

Atletiek

Beoordeling:	☆☆☆☆☆ 0 stemmen
Kosten:	nee
Leeftijdsgroep:	18+
Geplaatst op:	04 November 2012
Document type:	Internetpagina
Soort leermateriaal:	informatiebron
Eindgebruiker:	leerling/student
Grootte:	niveau 2: serie elementen
Studiebelasting:	-

[Sla op als favoriet](#) »

[Gebruik in arrangement](#) »

[Metadata aanvullen](#) »

Omschrijving

Leerlijnen

0 beoordelingen

Trefwoorden

[afstanden](#), [atleten](#), [atletiek](#), [circuits](#), [iaaf](#), [jeugd](#), [kampioenschappen](#), [olympische](#), [onderdelen](#), [plaats](#), [spelen](#), [sport](#), [verjaardag](#), [wedstrijden](#)

Omschrijving

Atletiek is een oude sport waar de sporters (atleten) individueel of in groepen (estafette) moeten presteren. Atletiek wordt zowel op de weg als op een atletiekbaan beoefend. Atletiek wordt wel 'de moeder der sporten' genoemd omdat het de menselijke basisbewegingen (lopen, springen, werpen) omvat.



lerniveau

Tekst Leesbaarheid

Flesch-Kincaid
Coleman-Liau
Dale-Chall
SMOG
KPC (oude AVI)

	words	FK	CL	DC	SMG	KPC
David Hamilton	505	17	21	21	19	13
Falerii Veteres	4173	20	21	21	21	15
Houten waal	770	15	18	21	17	12
Kerncentrale Tsjernobyl	1211	18	20	21	20	14
Linksnationalisme	635	23	26	21	21	17
Reduced instruction set computer	1047	19	23	21	21	15
ROOD, jong in de SP	1071	19	21	22	20	14
Shirley MacLaine	664	16	17	23	18	13
Viggo Mortensen	557	16	18	20	18	13
Willem George August Lochmann van Bennekom	627	19	18	22	21	14

Frequentie

- 1.299.478 unieke woorden
- 636.045 woorden komen 1x voor
- 1. “de” 10.333.955
- 2. “van” 5.333.294
- 100. “alle” 96.791
- 1000. “juist” 9.743
- 10000. “diefstal” 643



Serie elementen | **Bron:** [Wikipedia](#)

Lijst van atleten

Beoordeling: ★★★★★ 0 stemmen

Kosten: nee

Leeftijdsgroep: 1307+

Geplaatst op: 04 November 2012

Document type: Internetpagina

Soort leermateriaal: informatiebron

Eindgebruiker: leerling/student

Grootte: niveau 2: serie elementen

Studiebelasting: -

[Sla op als favoriet »](#)

[Gebruik in arrangement »](#)

[Metadata aanvullen »](#)

Omschrijving

Leerlijnen

0 beoordelingen

Trefwoorden

[david](#), [jan](#), [john](#), [olga](#), [paul](#), [robert](#)

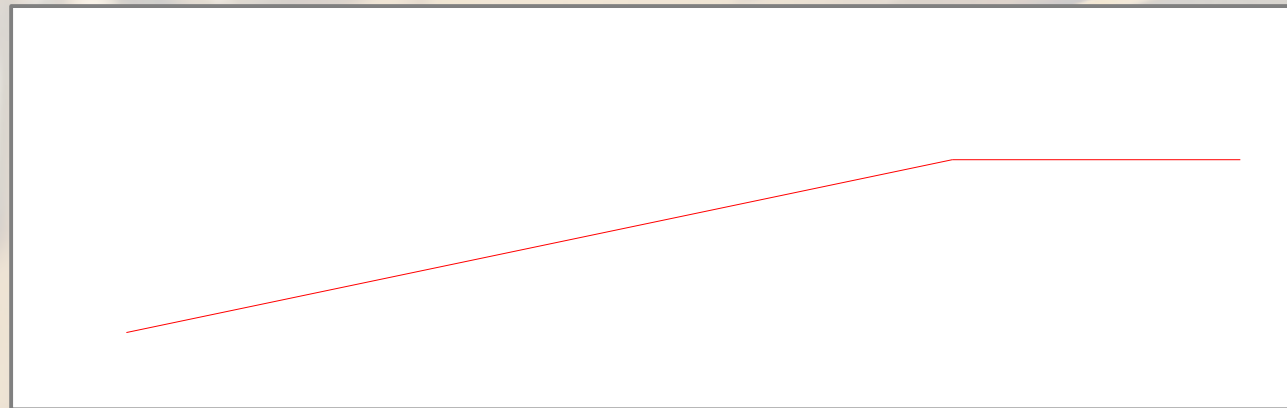
Omschrijving

Dit is een lijst van atleten.



studielast

wpm



leeftijd

x aantal woorden = studielast



Serie elementen | **Bron:** [Wikipedia](#)

De gouden vogel

Beoordeling:	★★★★★ 0 stemmen
Kosten:	nee
Leeftijdsgroep:	12+
Geplaatst op:	04 November 2012
Document type:	Internetpagina
Soort leermateriaal:	informatiebron
Eindgebruiker:	leerling/student
Grootte:	niveau 2: serie elementen
Studiebelasting:	12 minuten

[Sla op als favoriet](#) »

[Gebruik in arrangement](#) »

[Metadata aanvullen](#) »

Omschrijving

Leerlijnen

0 beoordelingen

Trefwoorden

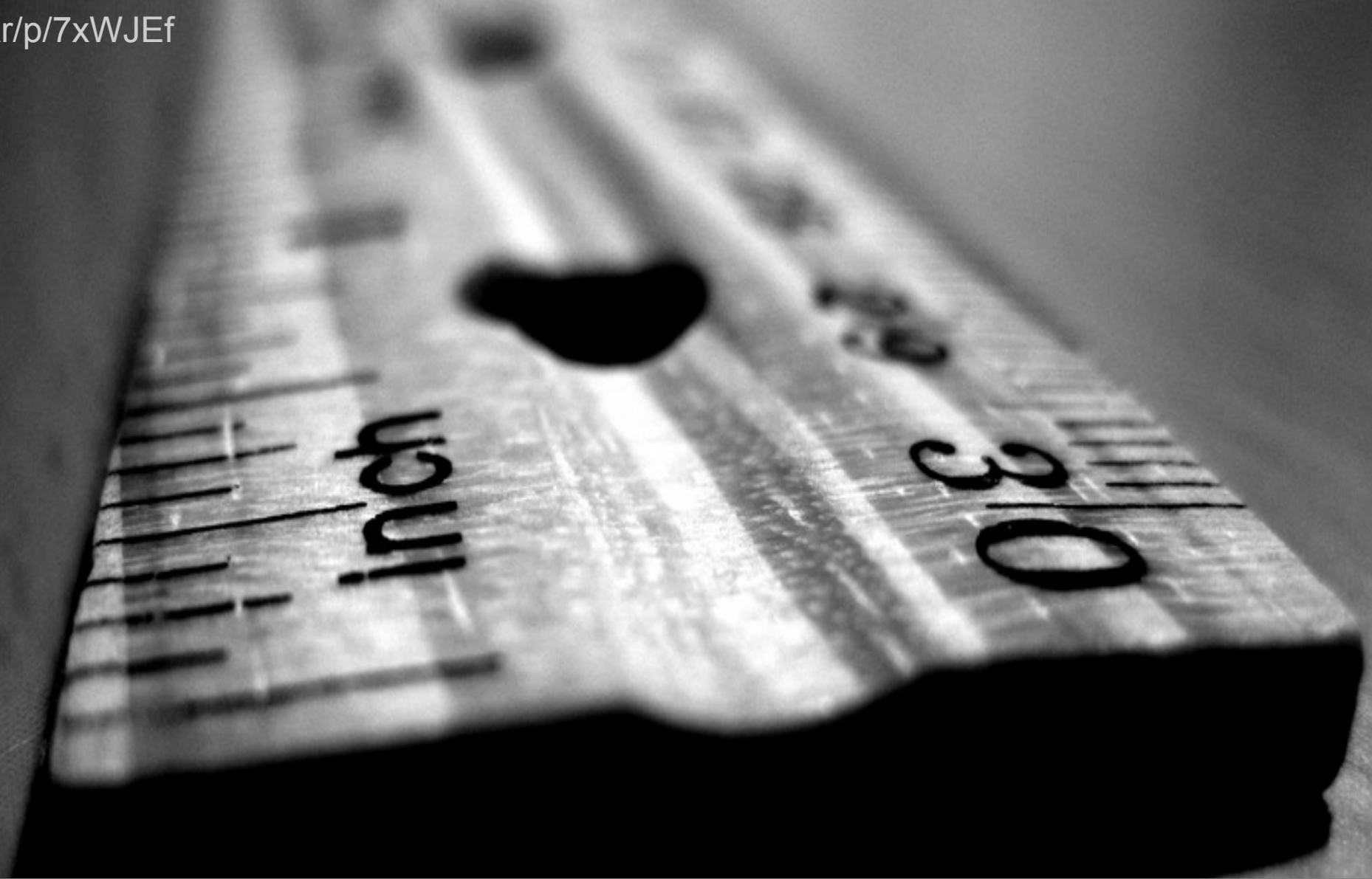
[appèl](#), [bos](#), [broers](#), [dag](#), [dorp](#), [gouden](#), [jongen](#), [jonkvrouw](#), [kasteel](#), [koning](#), [koningsdochter](#), [koningszoon](#), [kooi](#), [opnieuw](#), [paard](#), [raad](#), [vogel](#), [volgende](#), [weer](#), [ziet](#), [zoon](#)

Omschrijving

De gouden vogel is een sprookje, opgetekend door de gebroeders Grimm in Kinder- und Hausmärchen onder nummer KHM57. De oorspronkelijke naam is "Der goldene Vogel".



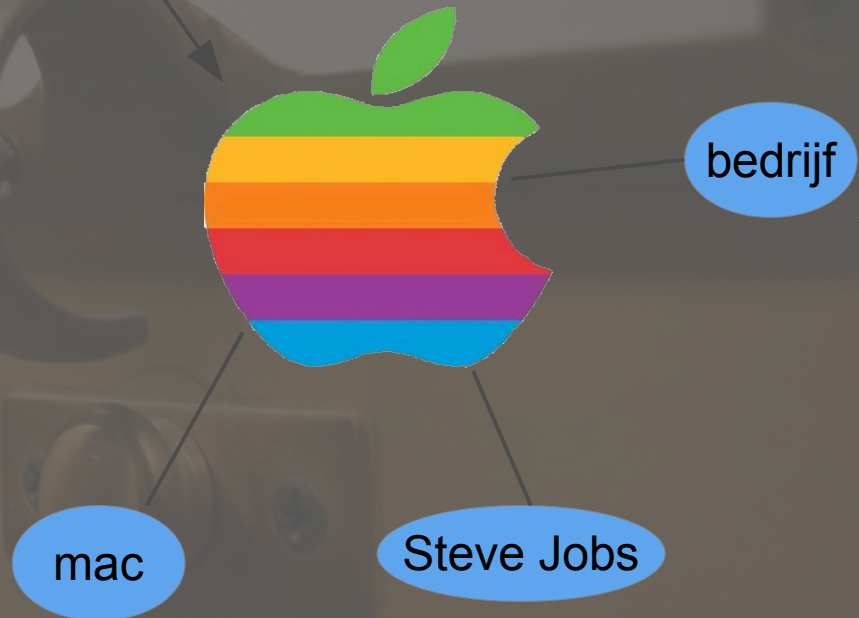
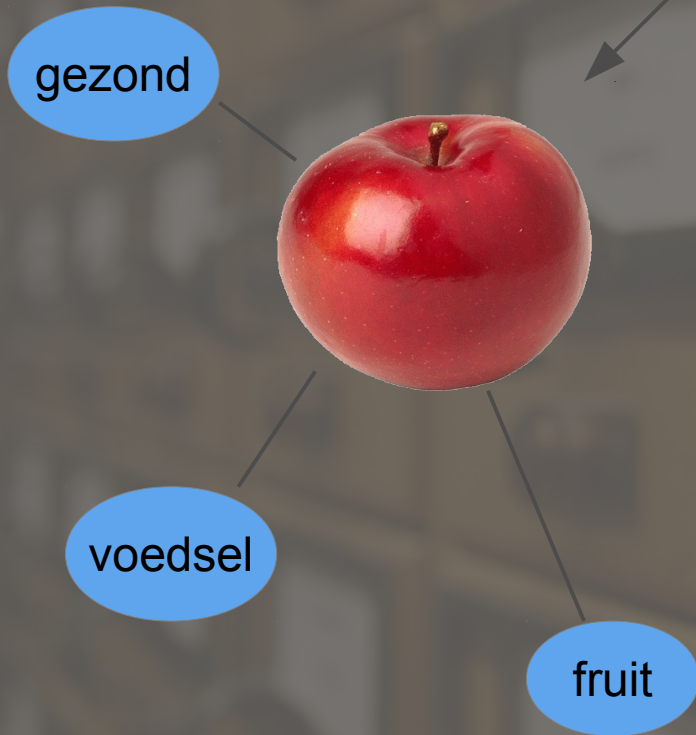
vakgebied

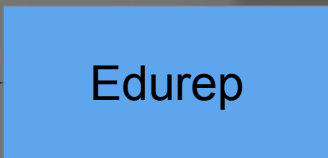
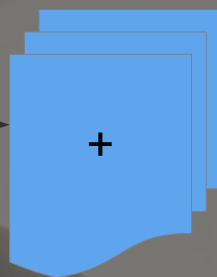


Mapping

Wikipedia categoriën naar onderwijs begrippen

Naive Bayes Classificatie





laatste slide

software:

NLTK, WikiExtractor, python-hyphenator, Myspell
dictionary van opentaal.org, readability-score

Wim Muskee
wimmuskee@gmail.com
www.flowledge.nl