XML

# Inleiding

**XML (eXtensible Markup Language):** groep regels voor het definiëren van semantische labels die een document in stukjes opdelen en de verschillende stukjes identificeren

* is de standaard voor het definiëren van markup-talen
* wordt gebruikt voor de representatie van gegevens (≠ opmaak)
* is zowel leesbaar voor mens als computer(taal)
* is geschreven in Unicode → meertalig
* is platformonafhankelijk (gegevens uitwisselen tussen applicaties op versch. platformen)
* is een metataal (taal met mogelijkheid zelf labels te maken)
* deze labels worden vastgelegd in een schema

(woordenlijst en syntaxis voor bepaalde documenten)

# Voorbeelden

* Chemical Markup Language (beschrijving van moleculen bestaande uit atomen en hun verbindingen)
* Mathematical Markup Language (beschrijven van wiskundige vergelijkingen)
* Really Simple Syndication (RSS)
* Scaleable Vector Graphics (SVG)

# Syntax

= zelfde als HTML, maar:

* elk XML-document begint met de **XML-declaratie**: <?xml version="1.0" standalone=”no"?>
  + start document met declaratie van versie
  + attribuut standalone
  + “yes” : document heeft geen andere bestanden nodig om te kunnen worden gelezen
  + “no” : bij gebruik van externe entiteiten en entiteitsverwijzingen (= standaardwaarde)
* is **veel strikter**
  + closing tags mogen nooit weggelaten worden (ook niet bij lege elementen)
  + alle XML-elementen moeten juist genest zijn

(opening en closing tag binnen dezelfde parent en op hetzelfde niveau)

* + XML-tags zijn hoofdlettergevoelig
  + attribuutwaarden moeten altijd tussen quotes staan
  + **<**, **>**, **&**, **‘** en **“** zijn verboden en moeten vervangen worden door hun *entity reference*
* bij tekstblokken met veel verboden tekens → CDATA-secties: <![CDATA[*inhoud*]]>
  + er mag geen overbodige *white space* zijn (spaties/tabs)
* XML-documenten moeten altijd een **root-element** hebben (een *parent* van alle andere elementen)

<root>  
  <child>  
    <subchild>.....</subchild>  
  </child>  
</root>

* namen van tags mogen **zelf gekozen** worden