

Wikipedia / WikiCon 2014

# Taxoboxen für Lebewesen: Bot oder Wikidata?

Oktober 2014

- Grundbegriffe
  - Taxobox, Bots, Wikidata
- Taxobot
  - Historie, Funktionsweise, derzeitige Abdeckung
- Taxonomie in Wikidata
  - Relevante Properties, Strategie
- Bot oder Wikidata?
  - Vergleich, Kombination möglich?

- Taxon (Plural Taxa)
  - Wikipedia: „Als Taxon bezeichnet man in der Biologie eine als systematische Einheit erkannte Gruppe von Lebewesen.“  
...  
„Die Einordnung nach dem Grad der Verwandtschaft in das entsprechende hierarchische System ist die Aufgabe der akademischen Disziplin der Taxonomie, die sich wiederum auf die Ergebnisse der biologischen Systematik stützt.“
  - Bioinformatik: Knoten im phylogenetischen Baum

- Rang
  - Hierarchische Stufe im taxonomischen Baum
  - Beispiele: Klasse, Ordnung, Familie, Gattung, Art ([Dok. Taxobox](#))
  - Ähnliche Standards in Zoologie (ICZN) und Botanik bzw. Mykologie (ICN), aber nicht einheitlich
  - Anmerkung
    - Bedeutung der traditionellen Ränge schwindet, klassische vs. kladistische Taxonomie
    - Aber: Orientierungshilfe

## ■ Taxobox

- Infobox mit übergeordneten Taxa zu einem Taxon
- Zudem Autor (Erstbeschreiber) zum Artikelaxon
- Variante: Paläobox (Dinosaurier)
  - gleiche Vorlage
  - zusätzliche Angaben wie Zeit oder Fundorte
- Parameter für übergeordnete Taxa ( $2 \leq x \leq 6$ )
  - | Taxon $_x$ \_Name =
  - | Taxon $_x$ \_WissName =
  - | Taxon $_x$ \_LinkName =
  - | Taxon $_x$ \_Rang =
  - | Taxon $_x$ \_Autor =

| Blaumeise   |                                |
|---|--------------------------------|
|  |                                |
| Blaumeise ( <i>Cyanistes caeruleus</i> )  |                                |
| Systematik  |                                |
| Klasse:   | Vögel (Aves)                   |
| Ordnung:  | Sperlingsvögel (Passeriformes) |
| Unterordnung:   | Singvögel (Passeri)            |
| Familie:  | Meisen (Paridae)               |
| Gattung:  | <i>Cyanistes</i>               |
| Art:  | Blaumeise                      |
| Wissenschaftlicher Name   |                                |
| <i>Cyanistes caeruleus</i><br>LINNAEUS, 1758  |                                |



6

⋮

2

- **Monotypische Taxa**
  - Taxon, das nur genau ein untergeordnetes Taxon enthält (Bsp. Ginkgo)
  - Konvention in der deutschsprachigen Wikipedia: Nur ein Artikel, Lemma entspricht niedrigerem Taxon
  - Probleme:
    - Autor des übergeordneten Taxons ist darzustellen
    - rezent / fossil
    - ...

| Systematik                          |                                    |
|-------------------------------------|------------------------------------|
| <i>Unterabteilung:</i>              | Samenpflanzen<br>(Spermatophytina) |
| <i>Klasse:</i>                      | Ginkgopflanzen (Ginkgoopsida)      |
| <i>Ordnung:</i>                     | Ginkgoales                         |
| <i>Familie:</i>                     | Ginkgogewächse                     |
| <i>Gattung:</i>                     | Ginkgo                             |
| <i>Art:</i>                         | Ginkgo                             |
| Wissenschaftlicher Name der Familie |                                    |
|                                     | Ginkgoaceae<br>Engl.               |
| Wissenschaftlicher Name der Gattung |                                    |
|                                     | <i>Ginkgo</i><br>L.                |
| Wissenschaftlicher Name der Art     |                                    |
|                                     | <i>Ginkgo biloba</i><br>L.         |

- Bots in der Wikipedia
  - Computerprogramm, das der automatisierten Bearbeitung oder Auswertung von Wikipedia-Seiten dient
  - Verschiedene Programmiersprachen
    - z.B. PHP, Java, Python
  - Verschiedene Frameworks
    - d.h. bereits implementierte Basisfunktionen
  - Umgebung
    - spezieller Account, Ausführung auf Server oder Client, periodisch oder bei Bedarf...

- Wikidata
  - zentrale Datenbank analog Commons
  - Motivation: Höhe eines Berges oder Einwohnerzahl einer Stadt ist in jeder Sprachversion dieselbe
  - zunächst (ab Oktober 2012) Aufnahme der Interwiki-Links
  - seit April 2013 Einbindung von Wikidata-Daten in deutschsprachiger Wikipedia möglich



| Systematik                             |                                     |
|--|-------------------------------------|
| Überfamilie:                           | Karpfenfischähnliche (Cyprinoidea)  |
| Familie:                               | Karpfenfische (Cyprinidae)          |
| Unterfamilie:                          | Cyprininae                          |
| Gattung:                               | <i>Carassius</i>                    |
| Art:                                   | Giebel ( <i>Carassius gibelio</i> ) |
| Form:                                  | Goldfisch                           |
| Wissenschaftlicher Name                |                                     |
| <i>Carassius gibelio forma auratus</i> |                                     |
| (BLOCH, 1782)                          |                                     |

```

{{Taxobox
| Taxon_Name           = Goldfisch
| Taxon_WissName      = Carassius gibelio forma auratus
| Taxon_Rang          = Form
| Taxon_Autor         = ([[Marcus Elieser Bloch|Bloch]], 1782)
| Taxon2_Name         = Giebel
| Taxon2_LinkName     = Giebel (Fisch)
| Taxon2_WissName     = Carassius gibelio
| Taxon2_Rang         = Art
| Taxon3_WissName     = Carassius
| Taxon3_Rang         = Gattung
| Taxon4_Name         =
| Taxon4_WissName     = Cyprininae
| Taxon4_Rang         = Unterfamilie
| Taxon5_Name         = Karpfenfische
| Taxon5_WissName     = Cyprinidae
| Taxon5_Rang         = Familie
| Taxon6_Name         = Karpfenfischähnliche
| Taxon6_WissName     = Cyprinoidea
| Taxon6_Rang         = Überfamilie
| Bild                = Goldfish3.jpg
| Bild

```

- Ausgangsproblem
  - Redundanz der übergeordnete Taxa in Taxoboxen
  - Konsistenz, Anpassungsaufwand bei Änderungen
  - Vorlage oder Bot?
- Grundgedanken
  - Zentrale Definitionsseite für Taxon-Baum als Wiki-Liste
  - Auch von Nicht-ITlern bearbeitbar oder erstellbar
  - Für klar abgrenzbare Teilbereiche einsetzbar
  - Detailgrad flexibel

- Technik
  - C# / Microsoft .NET
  - Verwendung des [DotNetWikiBot-Frameworks](#)
  - Zusätzliche weitere selbstimplementierte Basisfunktionalität
    - Vorlagen-Parser
    - Kategorie-Handling
    - ...
  - Ausführung
    - bei Bedarf (derzeit in etwa halbjährlich)
    - derzeit auf Client, nicht auf Wikimedia-Server

## ■ Historie

- Dezember 2009: Anpassung der Systematik an APG3 für die Pflanzen (Wiki Projekt)
  - Präferenz zunächst Vorlagen-basierte Lösung
  - Vorschlag Bot-basierte Lösung, Ausarbeitung der Details in Zusammenarbeit mit „Botanikern“, v.a. Griensteidl
- August 2011: Zweites Einsatzgebiet: Knochenfische
  - Ausgearbeitet durch Haplochromis
- Oktober 2012: Ergänzung Kategorieabgleich
- März 2013: Meldungen des Bot nun „öffentlich“

- Definitionsseite (Beispiel: Schlangen)
  - Vorspann
    - Quellenangabe
    - Ausgangskategorie
    - zu verwendende Statusseite
    - übergeordnete Taxa des Ausgangstaxons
    - Ausgangstaxon
  - Liste mit Taxobaum
    - pro Aufzählungszeile, getrennt durch Backslash
      - Rang
      - wissenschaftlicher Name
      - Trivialname
      - Linkname
      - Autor (nur nötig für monotypische Taxa)
      - ggf. Kommentar
  - Umbenennungen

- Funktionsprinzip
  - Abarbeiten aller Artikel zur Ausgangskategorie mit Taxobox
  - Durchsuchen der Taxobox von unten bis erstes bekanntes Taxon auftaucht. Kriterien:
    - Trivialname und Rang gleich
    - wissenschaftlicher Name gleich
  - Entsprechen dieses Taxon und die übergeordneten Taxa mit ihren Attributen der Definitionsseite?
    - Ggf. angleichen. Aber: Das Artikeltaxon wird nie angepasst.
    - [Beispieledit](#)

- Besonderheiten
  - Auch auszuschließende (Unter-)kategorien möglich
  - Name, WissName und LinkName können auf der Definitionsseite als Wiki-Links angegeben werden
  - Autor muss nur für monotypische Taxa angegeben werden (ggf. „unbekannt“)
  - In der oberen Hälfte der Box auszulassende Zwischentaxa sind durch „^“ zu kennzeichnen
  - „Inoffizielle“ Ränge sind durch „?“ zu kennzeichnen (z.B. „Serie“ und „Unterserie“ bei den Knochenfischen)

- Kategorieprüfungen ([...auslösende Anregung](#))
  - Kategoriebaum spiegelt den Taxon-Baum wider
  - Nur ausgewählte Ebenen werden als Kategorie abgebildet
  - In der deutschsprachigen Wikipedia paralleler Kategoriebaum für wissenschaftliche Namen und Trivialnamen
  - Auch Redirects für wissenschaftliche Namen sind zu kategorisieren
  - Automatisierte Korrektur von Kategoriezuordnungen nur möglich, wenn die Kategoriebaum nicht tiefer gegliedert ist als der Taxon-Baum
  - Problem Kategoriesortierung ([Beispieledit](#))

- Meldungsliste auf Statusseite ([Beispiel](#))
  - Attribute des Artikeltaxons widersprechen Definitionsseite
  - Entfernen von Zwischentaxa
  - Änderung von Autorenzitaten
  - Unbekannte Parameter in Taxobox
- Zusatzprüfungen
  - bisher nur Gültigkeit der Ränge, logisch korrekte Folge (z.B. Familie < Ordnung)

- Abdeckung
  - Stand September 2014
  - etwa 66 %

|                      |       |
|----------------------|-------|
| <b>Pflanzen</b>      | 12150 |
| Kakteen              | 1985  |
| Affodillgewächse     | 653   |
| Dickblattgewächse    | 343   |
| Hundsgiftgewächse    | 287   |
| <b>Knochenfische</b> | 2913  |
| Buntbarsche          | 257   |
| <b>Knorpelfische</b> | 430   |
| <b>Libellen</b>      | 343   |
| <b>Pilze</b>         | 2400  |
| <b>Käfer</b>         | 1121  |
| <b>Säugetiere</b>    | 3114  |
| Mäuseartige          | 466   |
| <b>Schlangen</b>     | 424   |
| <b>Amorphea</b>      | 141   |
| <b>Excavata</b>      | 57    |
| <b>Amphibien</b>     | 463   |
|                      | 27547 |

# Taxonomie in Wikidata



# WIKIDATA

**Python regius**



**systematics**<sup>[1][2]</sup>

|             |                          |
|-------------|--------------------------|
| infraorder  | Alethinophidia           |
| superfamily | Pythonoidea              |
| family      | Pythonidae               |
| genus       | python ( <i>Python</i> ) |
| species     | <i>Python regius</i>     |

**scientific name of species**

*Python regius*  
 SHAW, 1802

**red list status**

|         |            |    |    |    |               |    |
|---------|------------|----|----|----|---------------|----|
| Extinct | Threatened |    |    |    | Least Concern |    |
| EX      | EW         | CR | EN | VU | NT            | LC |

- Wikidata, Grundzüge des Datenmodells
  - Items („Datenobjekte“) (Bsp.: [Amsel](#))
    - Verknüpft mit Seiten in Wikipedia
    - Enthält Properties
    - Identifiziert mit Q-Nummer
  - Properties („Eigenschaften“)
    - Menge der möglichen Properties ist vorgegeben
    - Beispiel „taxon name“
    - Identifiziert mit P-Nummer
    - Property mit Wert nennt man ein Statement („Aussage“)
    - Reference („Quelle“) für jedes Statement möglich
    - Mehrere Statements pro Property möglich ([Rank](#))

- WikiProjekt Taxonomie ([Projektseite](#))
  - Initiiert im Mai 2013 durch FelixReimann und Succu
  - Derzeit 12 eingetragene Teilnehmer
  - Ziele
    - Aufstellung der Regeln für taxonomische Daten in Wikidata
    - Implementierung der Regeln
    - Erfassung der Daten
  - Prinzipien
    - Jedes Taxon ist ein Item
    - Auch historische Sichtweisen werden festgehalten ([Cactaceae](#))

- Relevante Properties
  - instance of: „Taxon“ oder Spezialisierung davon
  - taxon name: Wissenschaftlicher Name
  - taxon rank: Item zu Rang
  - taxon author: Item des Autors (Erstbeschreibers)
  - date of scientific description: Jahr der Erstbeschreibung
  - parent taxon: Item des übergeordneten Taxons

- Weitere Eigenschaften
  - Trivialname
    - Label in der entsprechenden Sprache
  - Erstbeschreiber
    - Kürzel und Link über dessen Item zu ermitteln
  - incertae sedis
    - Als Property unter „parent taxon“ mit Angabe des Rangs
  - Redundante Properties für bestimmte Rangstufen
    - wurden/werden eliminiert ([Löschdiskussion](#))

# Taxonomie in Wikidata

- Taxobox aus Wikidata
  - Prototyp: [Module:Taxobox](#)
- Technisches Problem
  - In der Wikipedia ist bisher nicht möglich, auf Items (und deren Properties) zuzugreifen, die nicht direkt mit der Seite (dem Artikel) verbunden sind
  - Es gibt einen [Bug](#) hierzu

**Rosen**



[systematics](#)<sup>[1][2][3][4][5][6][7][8][9]</sup>

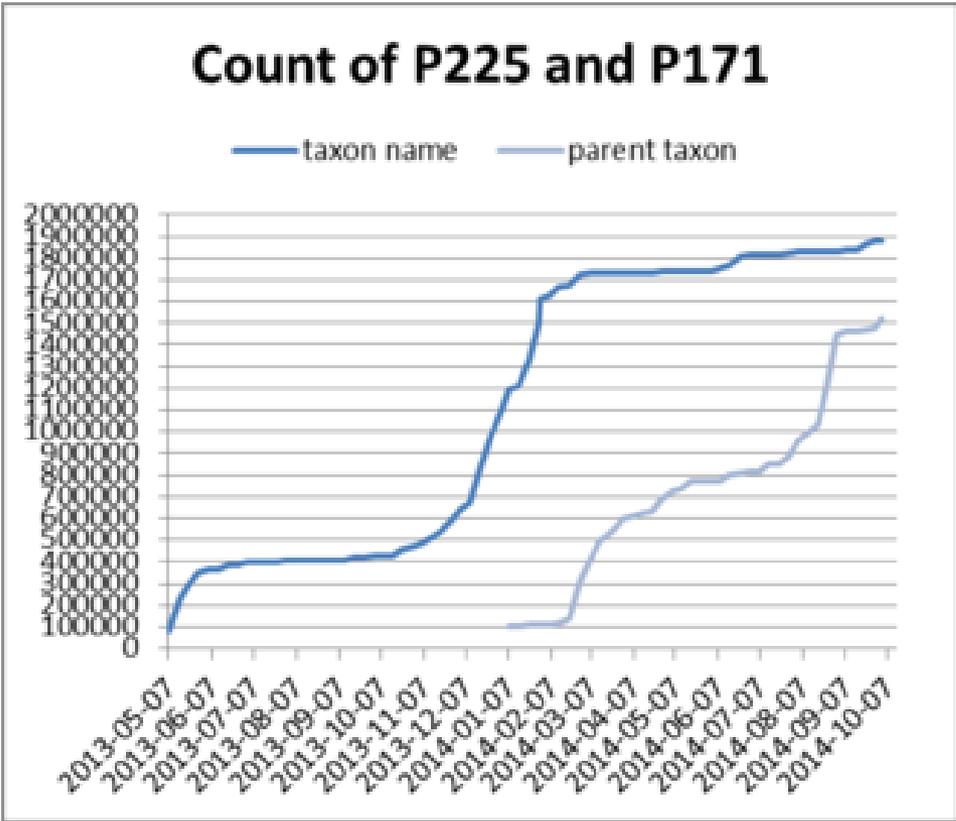
|              |                             |
|--------------|-----------------------------|
| Unterklasse  | Rosidae                     |
| Überordnung  | Rosanae                     |
| Ordnung      | Rosenartige<br>(Rosales)    |
| Familie      | Rosengewächse<br>(Rosaceae) |
| Unterfamilie | Rosoideae                   |
| Gattung      | <i>Rosa</i>                 |

**scientific name of Gattung**

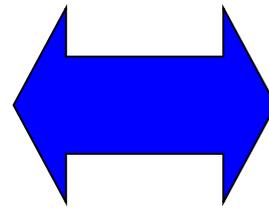
*Rosa*

# Taxonomie in Wikidata

- Abdeckung
  - Stand Oktober 2014



# Bot oder Wikidata?



- **Abgedeckter Umfang**
  - **Taxobot**
    - übergeordnete Taxa und deren Attribute
    - ausgewählte Teilbereiche der Systematik
    - nur deutschsprachige Wikipedia
    - auch Kategorien und Weiterleitungen für wissenschaftliche Namen
  - **Wikidata**
    - gesamte Systematik
    - alle Sprachen
    - auch Attribute des Artikeltaxons (Erstbeschreibung)

- Weitere Unterschiede
  - Art der Definition
    - Taxobot: auf einer Seite
    - Wikidata: Properties (insbesondere „parent taxon“) verteilt über Items
  - Bildung von Teilbereichen
    - Taxobot: Geltungsbereich einer Definitionsseite über Kategorien festlegbar
    - Wikidata: Bislang kein Verfahren
  - Anpassung der Artikel bei Änderungen
    - Taxobot: Edit nötig, dadurch aber auf Beobachtungsliste sichtbar
    - Wikidata: kein Edit im Artikel nötig, aber wie sicherzustellen, dass Artikeltext konsistent zu diesen Änderungen ist?

# Bot oder Wikidata?

- Kombination beider Ansätze möglich?
  - „Bewusste“ redundante Properties?
    - Klares Verfahren, solche Properties zu kennzeichnen
      - evtl. zunächst nur durch Konvention (standardisierte Namen)
      - aber am besten technisch unterstützt
      - nur durch Bots bearbeitbar
      - standardmäßig nicht sichtbar
  - Definition über zentrale Seite für Teilbereiche?
    - dann aber ist die Property „parent taxon“ nur für solche Taxa redundant, die durch eine solche Seite abgedeckt sind
    - Bisherige Ansätze: [\[1\]](#) [\[2\]](#) [\[3\]](#)

# Danke...

- ...fürs Interesse
- ... und Zuhören

