Internal Links in Wikipedia

Why are they not great and how should we fix them

A fantasy in three parts and two cats

Amir E. Aharoni
Wikimania 2019
0: Prologue
And all the things
That we spoke among us
Began with the words,
“But we”.

(Yehuda Amichai)
1: Red links
Obvious stuff: link

[[Whale]]
Obvious stuff: trail

[[Whale]]s
"Moby Dick" is a 1998 miniseries based on the 1851 novel of the same name.
Stop.

[[Did you begin editing many years ago|Is it actually obvious]]?
"The Cakemaker" is a drama film that stars [[Tim Kalkhof]].
Red in Hebrew, blue in German

"היאופה במברלין', הוא סרט בכיכובו של
טים קלקרוף." [spiel Kleegehr]
"The Cakemaker" is a drama film that stars [[Q19664702|Tim Kalkhof]].
"האופנה מברלין" הוא סרט בכיכובו שלטים קלקהוף.

"האופנה מברלין" הוא סרט בכיכובו שלטים קלקהוף.

Solution?
Awful?

\_\_ (ツ) _/\_
Not awful?

- [[Link]]s and page titles are text.
- In 2001 [[Tim Kalkhof]] was cooler than <a href="Tim_Kalkhof">Tim Kalkhof</a>.
- Is this still cool in 2019?
Not awful?

- In 2019 we know much more about maintaining a massively-multilingual encyclopedia.
- In 2019 we know much more about linked data.
- Links and titles are \textit{text}, but they point to \textit{things}. 
Not awful?

- If editing is mostly visual, then it doesn’t matter that you don’t have text as the target.
- Perhaps you can link the empty page itself instead of using the Q number for the link, but why bother.
What do we get?

- Automatic fallback: If there’s no article in English, you get a *chance* to read in German. Better than nothing.
- Automatic fallback: If there’s no article in a language we can read, we can see in Wikidata that it’s a German male actor, born in 1987 in Heidelberg. Also better than nothing.
What do we get?

- Consistency of ontology across languages
- Consistency of spelling of red links
- Automatic disambiguation →
2: Disambiguation
Disambiguation

''Felix'' may refer to:
* [[Felix (name)]], people and fictional characters with the name

==Places==
* [[Felix, Spain]], a municipality of the province Almería, in the autonomous community of Andalusia, Spain.
* [[Felix, Ontario]], an unincorporated place and railway point in Northeastern Ontario, Canada

== See also ==
* [[San Félix (disambiguation)]], the name of two places in South America
  * {{lookfrom}}

{{disambiguation}}
Disambiguation
Disambiguation

This slide was intentionally left blank
Imagine a world in which every disambiguation page can write itself.
But can they write themselves?

• Yes!
But can they write themselves?

• Yes!
• Computers can find identical titles.
• Humans have better things to do.
• Computers can usually find similar titles, too.
• In the worst case, humans can help computers find similar titles with Wikidata properties.
But can they write themselves?

• {{wiktionary|felix|Felix|Félix|FELIX}}

• Wikidata Lexemes can auto-create this.
• Humans have better things to do.
But can they write themselves?

- {{lookfrom}}
  - → All pages with titles beginning with Felix
  - Humans have much better things to do!
  - Software can auto-create this.
But can they write themselves?

- Humans have much, much better things to do!
- Why do we even care about this when instead, we could imagine a world in which...
Football (disambiguation)

Football is a family of sports that involve kicking a ball with the foot to score a goal.

Football may also refer to:

Sports

- Association football (known as football in most of the world and soccer in North America and Oceania)
- Australian rules football
- Gaelic football
- Gridiron football or football in North America and gridiron in Australia and New Zealand
  - American football
  - Canadian football
- International rules football
- Rugby football
  - Rugby league
  - Rugby union
- Football (ball), any of several types of balls used to play football (disambiguation)
Football (disambiguation)

Look up football in Wiktionary, the free dictionary.

Football is a family of sports that involve kicking a ball with the foot to score a goal.

Football may also refer to:

^ Sports

- Association football (known as football in most of the world and soccer in North America and Oceania)
- Australian rules football
- Gaelic football
- Gridiron football or football in North America and gridiron in Australia and New Zealand
  - American football
  - Canadian football
- International rules football
- Rugby football
  - Rugby league
  - Rugby union
- Football (ball), any of several types of balls used to play different sports
3: Titles
What is better?

C#[^b] (pronounced C sharp) is a general-purpose, multi-paradigm programming language encompassing strong typing, lexically scoped, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines.[^17] It was developed around 2000 by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2018). C# is one of the programming languages designed for the Common Language Infrastructure.

C#[^b] (pronounced C sharp) is a general-purpose, multi-paradigm programming language encompassing strong typing, lexically scoped, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines.[^17] It was developed around 2000 by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2018). C# is one of the programming languages designed for the Common Language Infrastructure.
What is worse?
Are these disambiguations?

- Voodoo Child (Slight Return)
- Add N to (X)
- Fear (Of the Unknown) (Grey's Anatomy)
¿Melocotón? ¿Durazno?
Prunus persica
especie de árbol


Prunus persica, originalmente Amygdalus persica L., melocotonero (del latín malus cotonus, «manzana algodonosa» —en alusión a la piel del fruto—)[3] es una especie de árbol del género Prunus de la familia Rosaceae. También se suele denominar duraznero (del latín durus acinus, «que tiene la piel dura» —aludiendo a la piel del fruto—) y a su fruto, durazno en partes de Canarias (en la actualidad melocotón es bastante usado) y en países como Argentina, Bolivia, Chile, Colombia, Costa Rica, Guatemala, El Salvador[4], Ecuador, México, Paraguay, Perú, Uruguay y Venezuela.
Can we deprecate MediaWiki titles?

• Wikidata labels are localizable:
  • English: Autumn / Fall
  • Spanish: Durazno / Melocotón
  • German: Januar / Jänner
Can we deprecate MediaWiki titles?

• Let’s replace most MediaWiki content page titles with Wikidata labels
• Wikidata labels have no limitations of characters
  • citation needed
• Wikidata description was always intended for disambiguation
∞: Epilogue
Wikimania

- Wikimania 2010 had a talk about the need to make interlanguage links centralized
- Wikimania 2011 had three talks about the need to make interlanguage links centralized
- In 2013 interlanguage links became centralized with Wikidata
Thanks / Argue with me