Wiki Migration
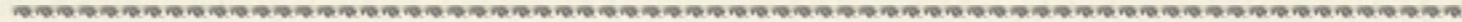

Dr. Vadim Zaytsev
(User:Spider, @grammarware)

Outline

Wiki
Migration

Dr. Vadim Zaytsev
(User:Spider, @grammarware)
©①①②②2011

Motivation

Y U MIGRATE

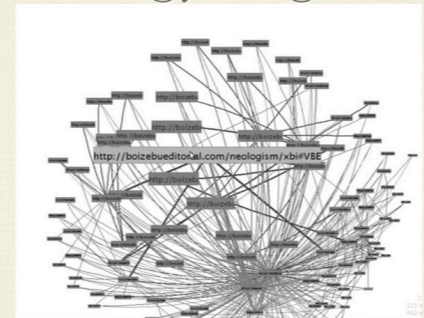7 slides

Community Migration

6 slides

Content Migration

9 slides

http://www.rascal-mpl.org/

Ontology Migration

6 slides

Thank you!

http://grammarware.net/

http://www.rascal-mpl.org/

Questions?

Motivation

7 slides

Wiki forking (community split)
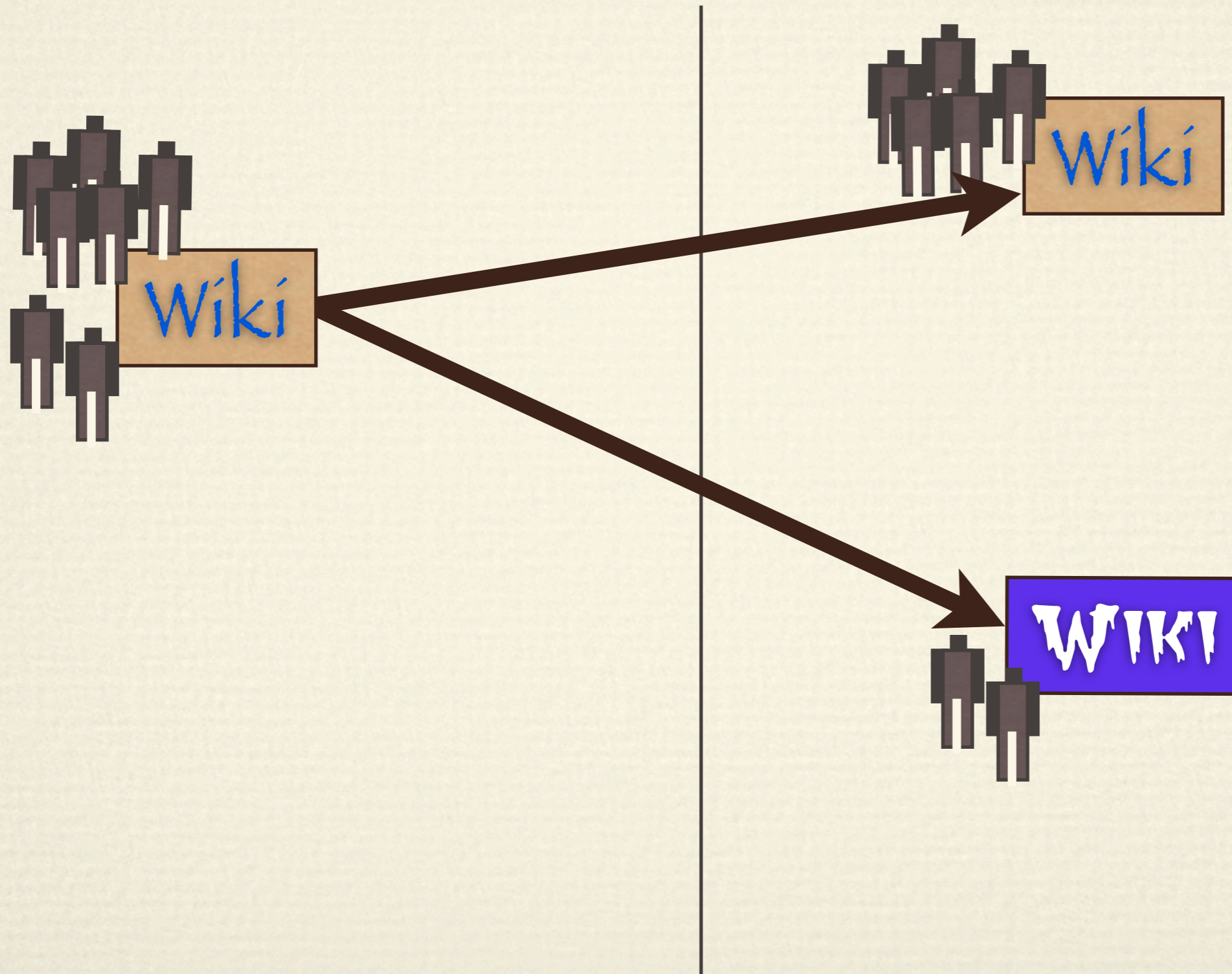
Wiki forking (content split)

Wiki merging (true merger)

Wiki merging (acquisition)

Wiki moving (upgrade)
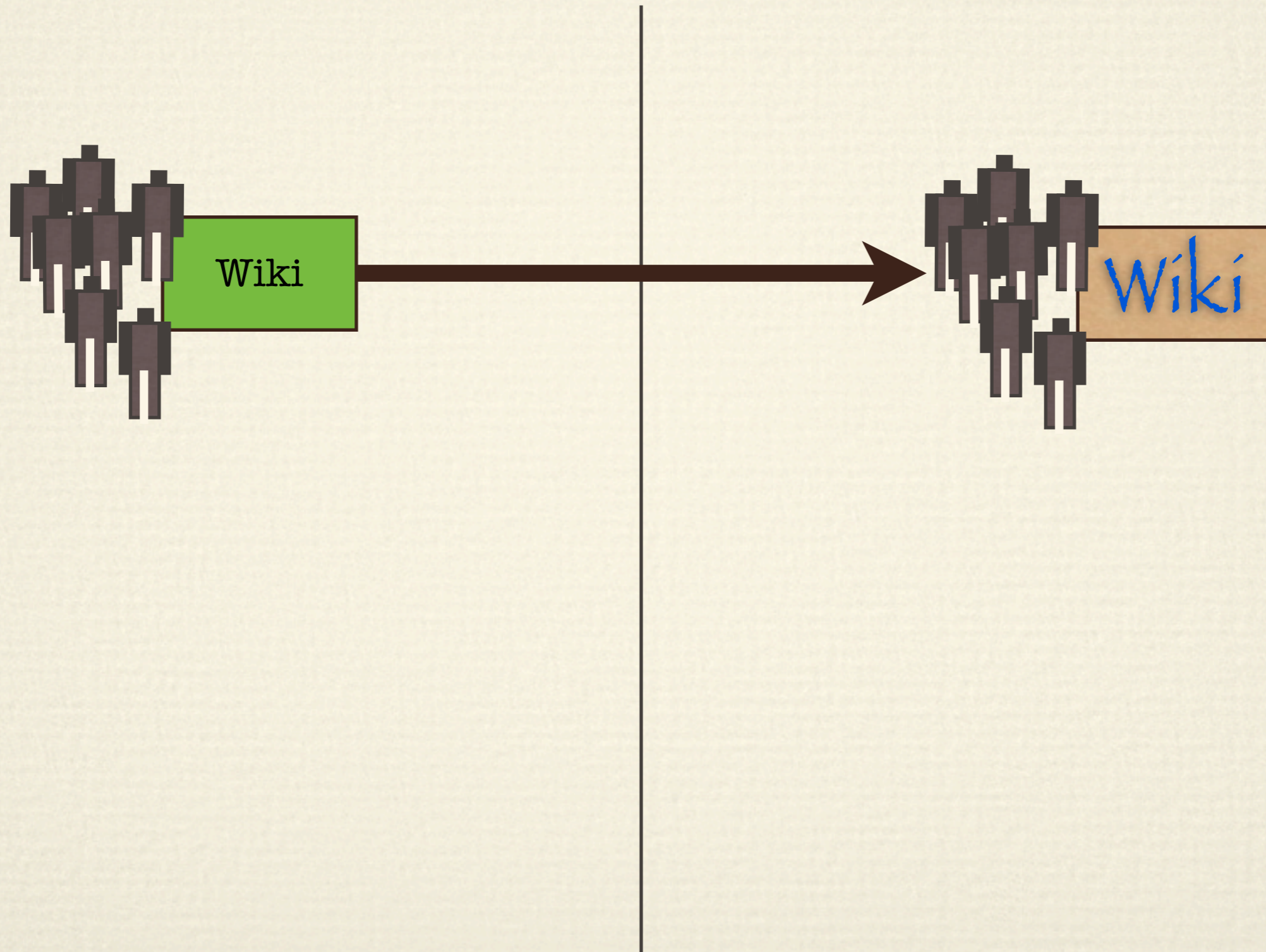
Wiki

Wiki

Wiki moving
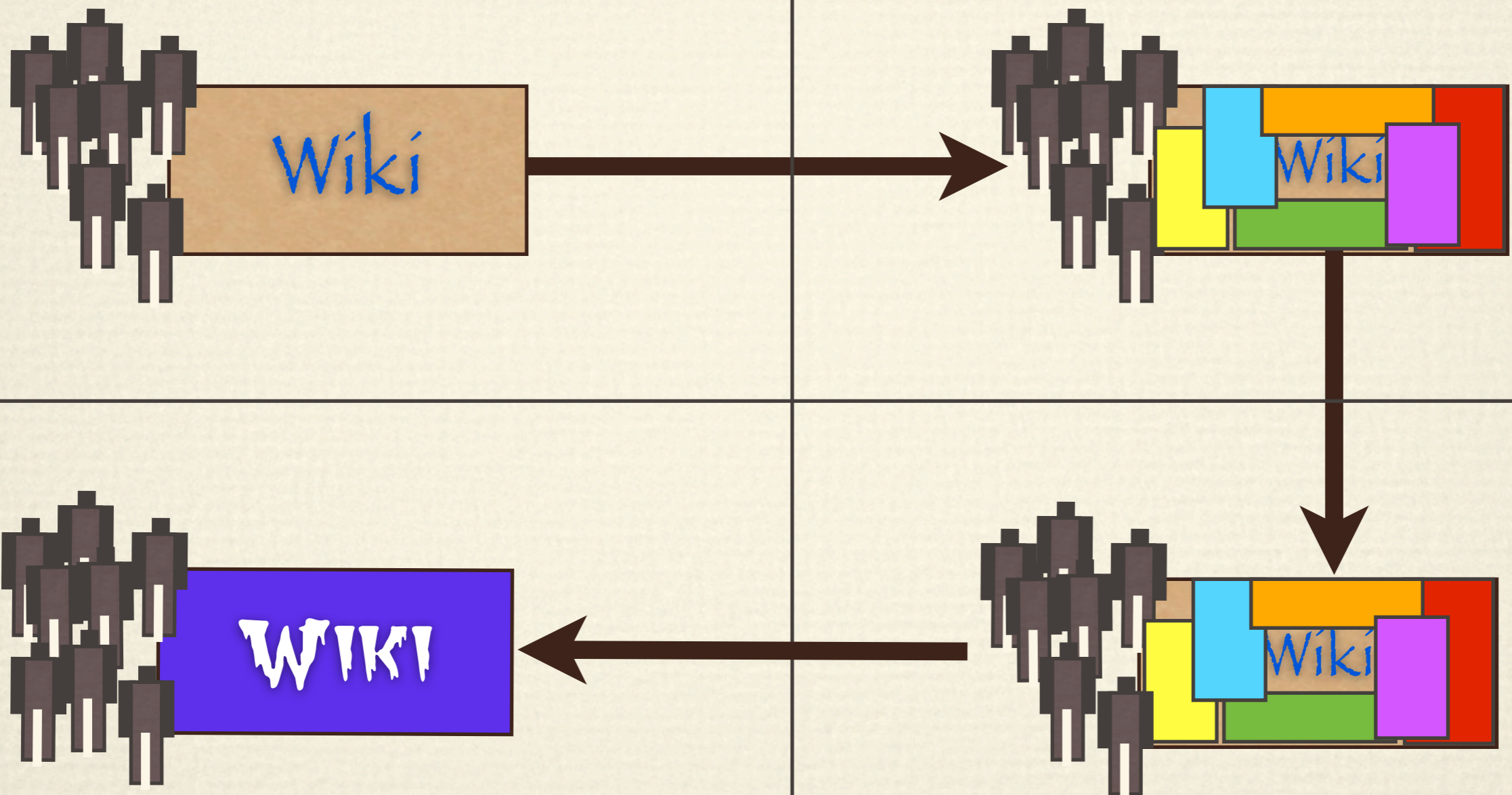
Alternative wiki front-end

Community Migration

6 slides

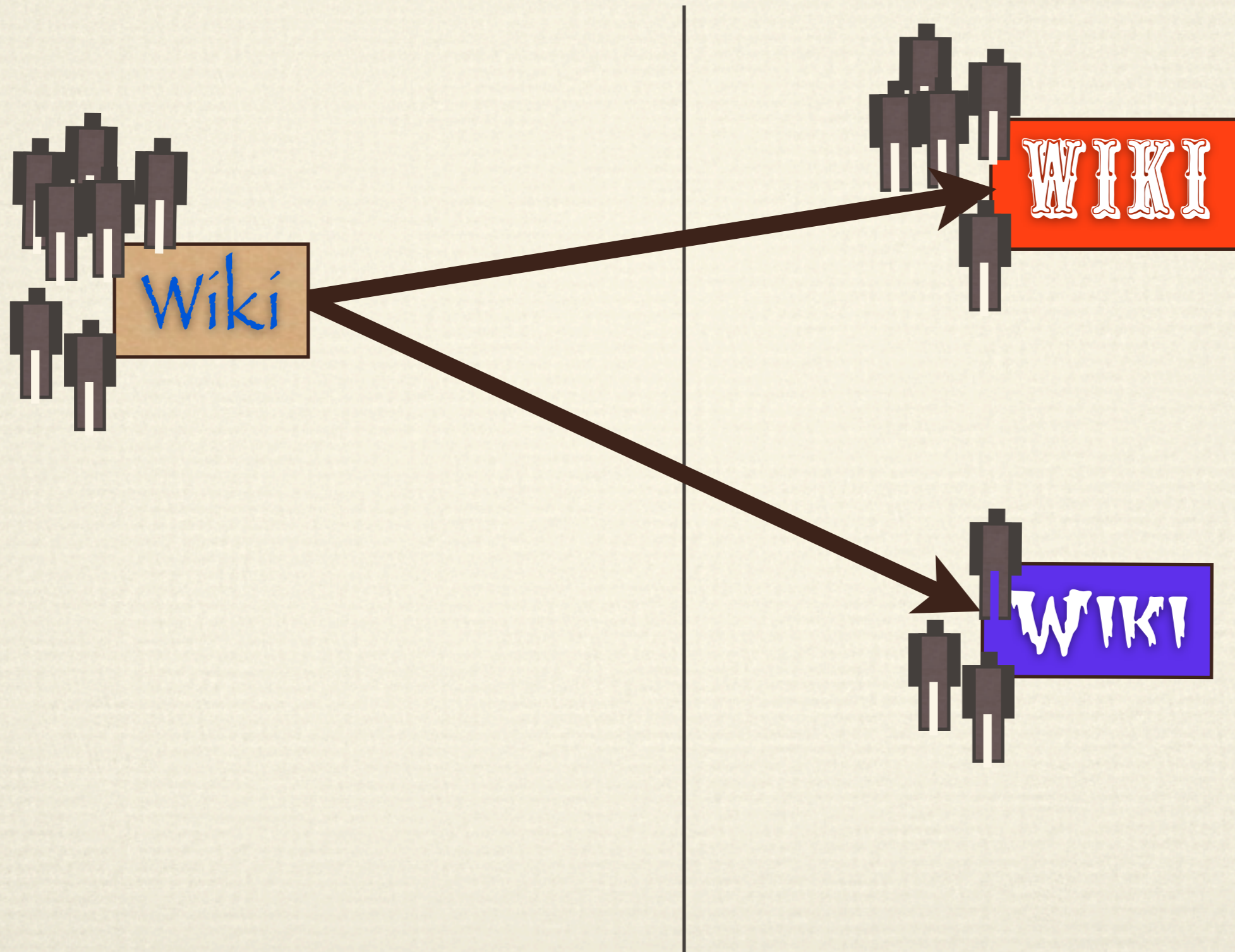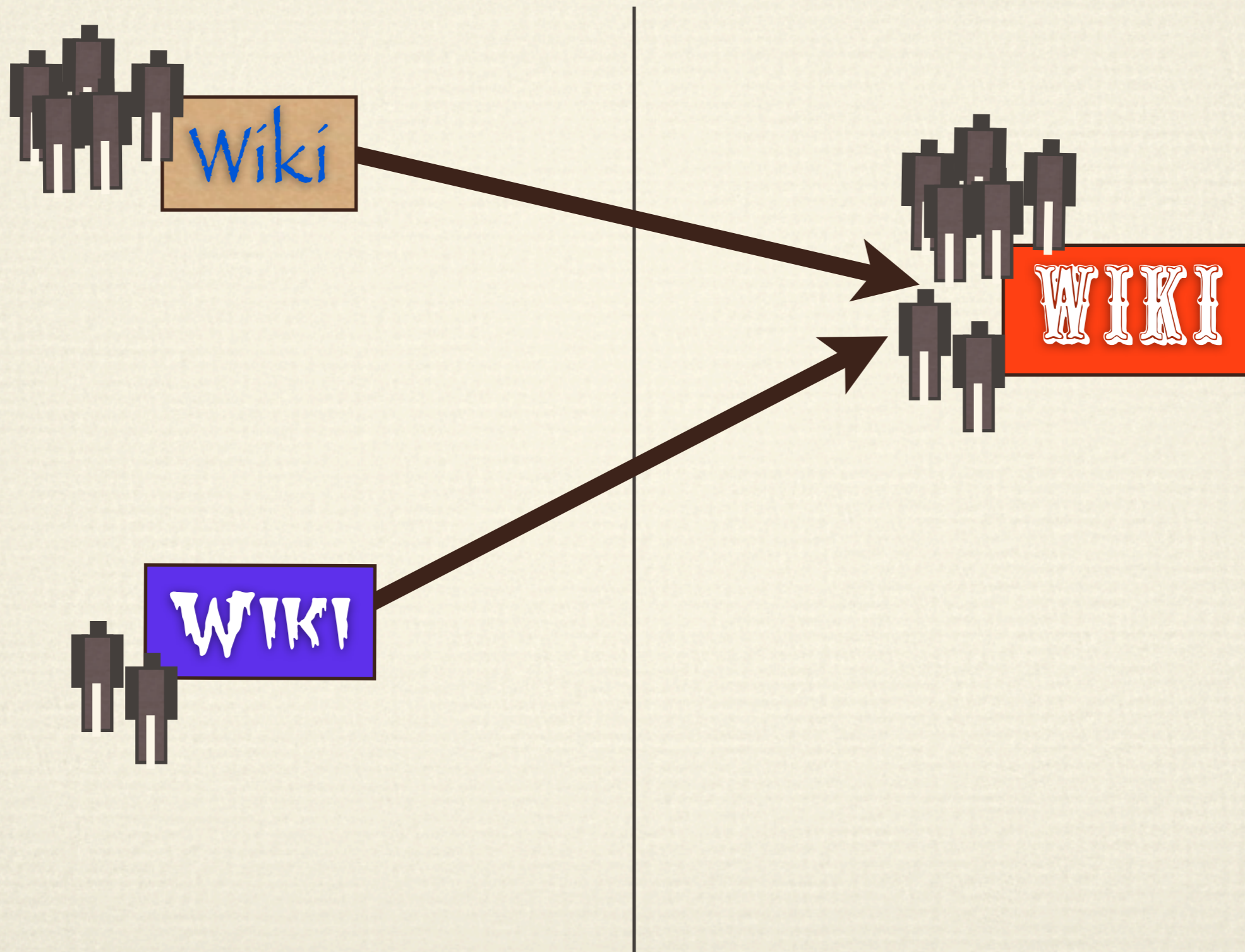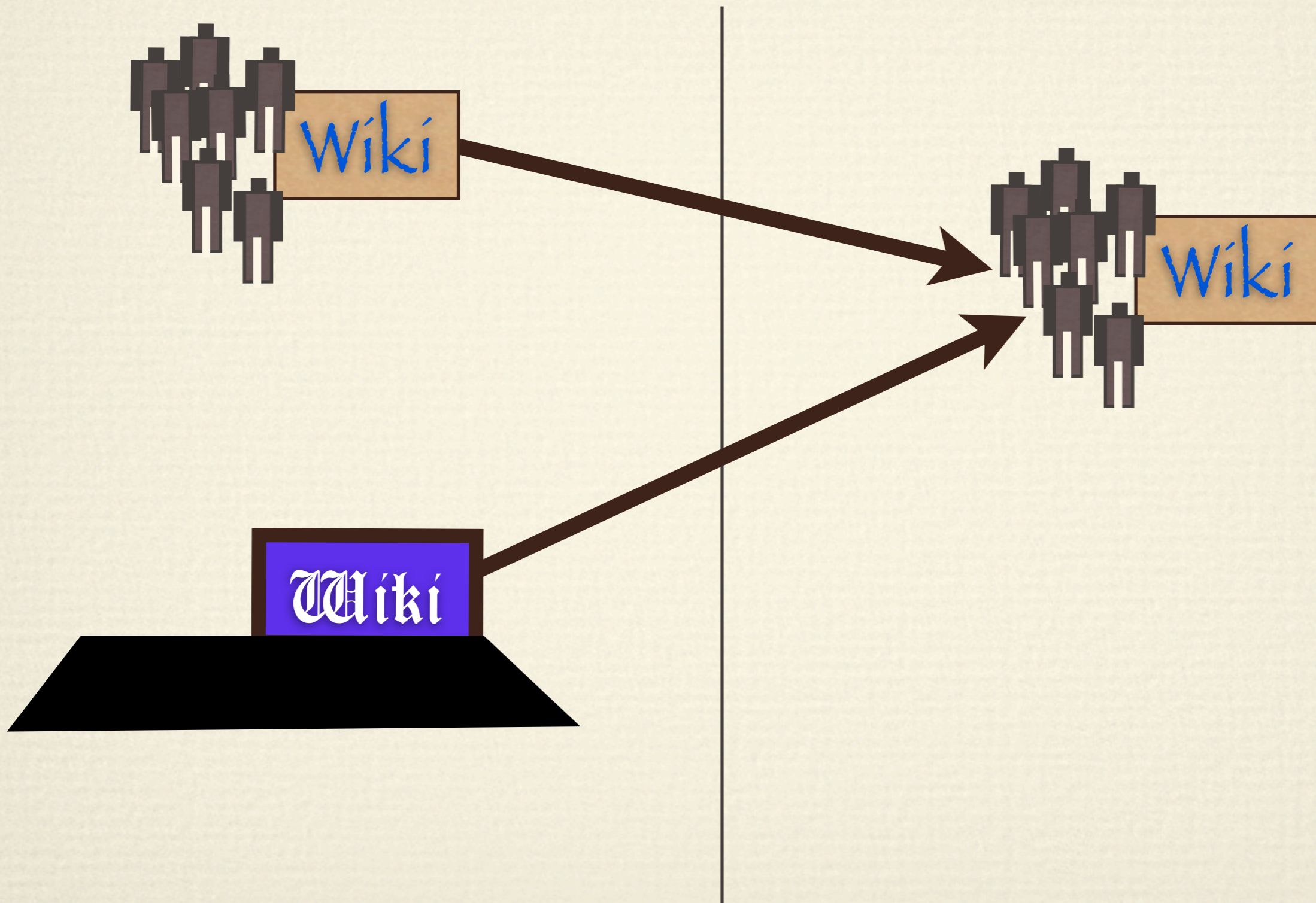*Persepolis* (2007), film by Marjane Satrapi and Vincent Paronnaud
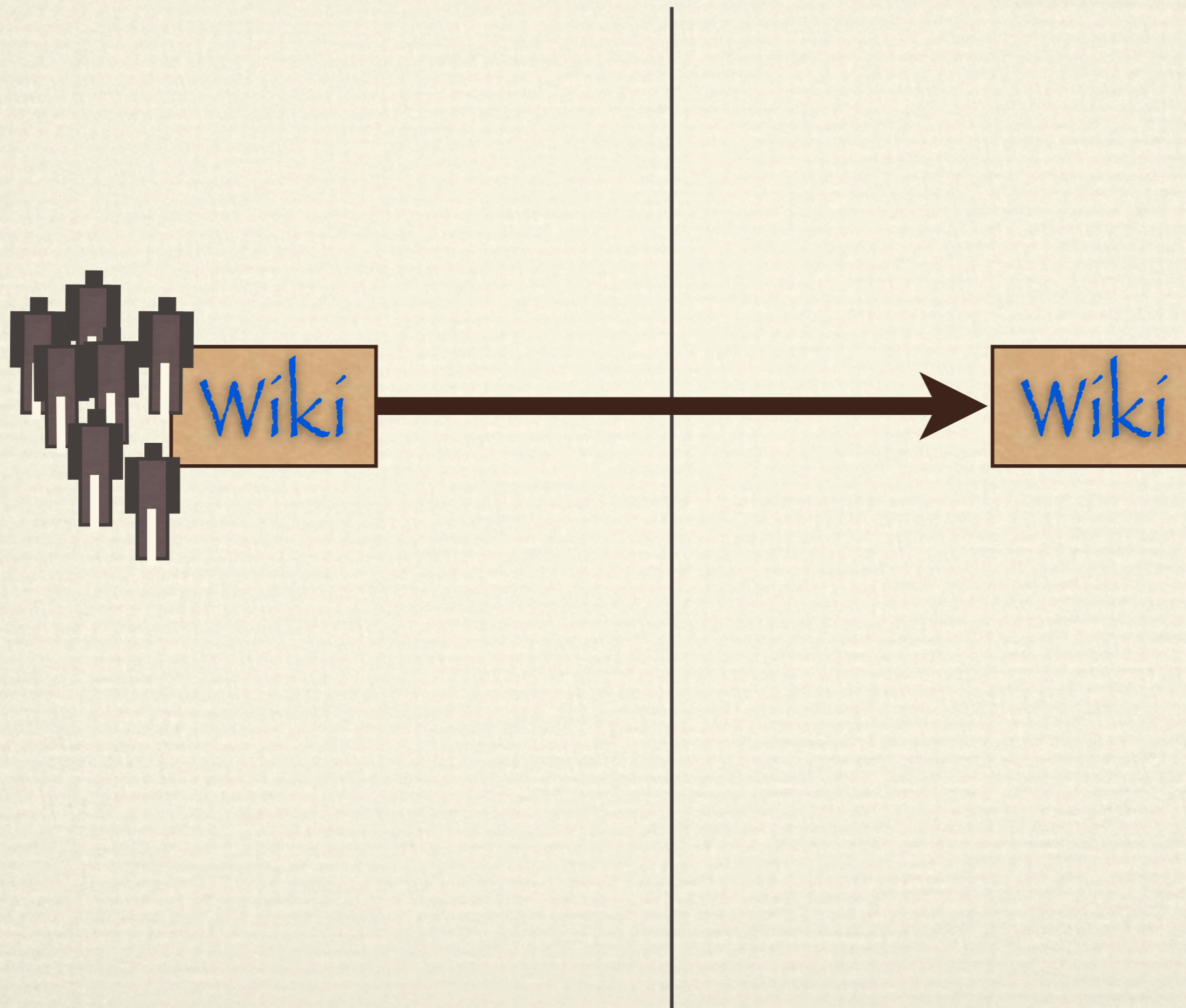
Wiki community

✓ System operators & bureaucrats

✓ Technical administrators (and wikignomes)

✓ Active content generators (exopedians)

✓ Active content beautificators (wikifaeries & wikiimps)

✓ Discussion junkies (metapedians)

✓ Supplementary content creators

✓ Novice wikipedians (wikipuppies)

✓ Dormant participants (wikiangels & wikidragons)

✓ Wiki content readers (wikiprincesses)

Community roles

✓ Admin != Admin

✓ Citizen, Editor, Officer (Citizendium)[2]

✓ SysOp, Bureaucrat, Check User, Editor (…)

✓ Warden, Expert, Master, Manager (WikiReality)[2]

✓ Expert, Master, Warden (CycloWiki)[2]

✓ SysOp, Bureaucrat, Check User, Editor, Reviewer, Moderator, Abuseditor (Lurkmore)[2]

✓ …

Migrating community roles (1/2)

✓ Technical staff

  ✓ needs to be re-educated

  ✓ documentation & links

✓ Exopedians

  ✓ provide maximum information

  ✓ be as helpful as you can

✓ WikiFaeries and WikiGnomes

  ✓ give time to embrace new place & concepts

Migrating community roles (2/2)

- ✓ Metapedians
  - ✓ value your time
  - ✓ many issues to discuss
- ✓ Illustrators etc
  - ✓ nice to have
  - ✓ don't be afraid to ask
- ✓ WikiPuppies
  - ✓ make it easy

Community migration

✓ Re-registration?

✓ User information migration

✓ User statistics migration

✓ Passwords migration (imply trust in the admins)

✓ Name clashes

✓ Namespace clashes

Broader picture

✓ Search engine optimisation

✓ Informing the target audience

✓ Static links to your wiki

✓ Inside the wiki: portals, to-do lists, etc

✓ Integration

✓ Friends & sisters

Content Migration

9 slides

Content migration

✓ Special:Export & Special:Import

  ✓ MediaWiki

✓ Wiki interchange format

  ✓ WS3, WIF, WAF, …

✓ Web scraping (HTML $\Rightarrow$ Wiki)

✓ Wiki dialect mapping (Wiki $\Rightarrow$ Wiki)

✓ Rendering and HTML importing

  ✓ TWiki, JSPWiki, …

Content migration methods

✓ Lexical approach

  ✓ one time effort

  ✓ very error prone

✓ Formal grammars

  ✓ good for programming languages

✓ Skeleton grammars

  ✓ more tolerant
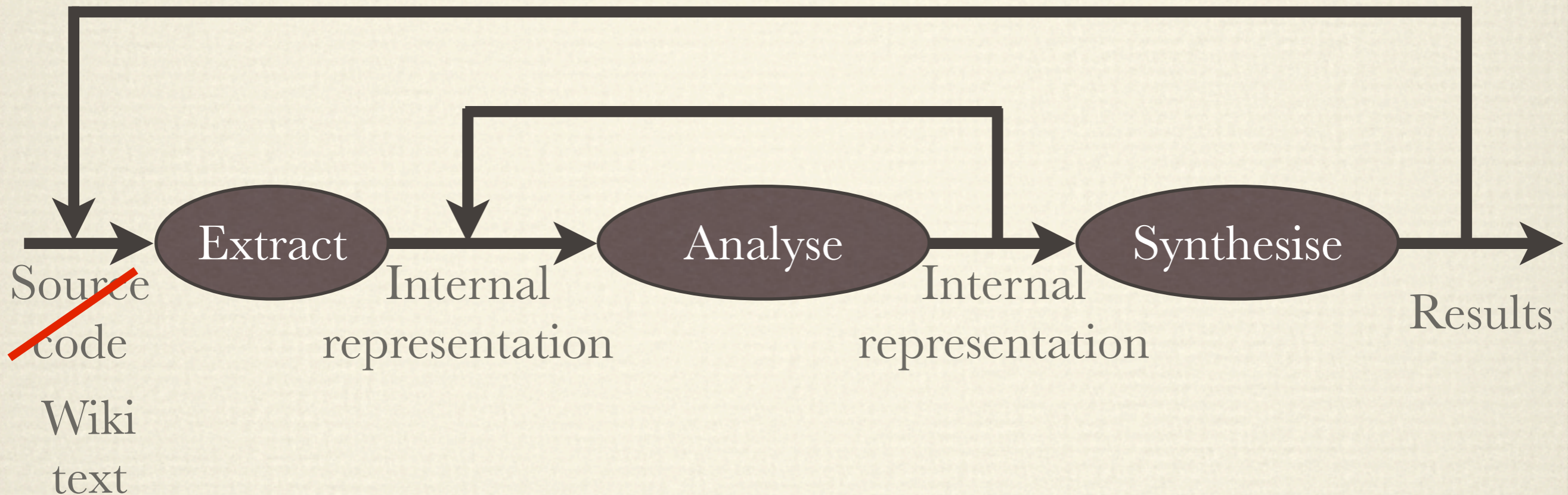
✓ Semi-automated manipulation of structured content

Wiki syntax: fonts/emphasis/...

	MediaWiki	TWiki	Wikidot	TiddlyWiki
Bold type	'''text'''	*text*	**text**	''text''
Italics	''text''	_text_	//text//	//text//
Bold italics	'''''text'''''	__text__	//**text**// **//text//**	//''text''// ''//text//''
Monospace	<tt>text</tt>	=text=	{{text}}	{{{text}}}
Monospace bold	<tt>'''text'''</tt> '''<tt>text</tt>'''	==text==	{{**text**}} **{{text}}**	{{{''text''}}} ''{{{text}}}''
Underline	<u>text</u>	<u>text</u>	__text__	__text__
Strikeout	<s>text</s>	<s>text</s>	--text--	--text--

Wiki syntax: links/images/…

	MediaWiki	TWiki	Wikidot	TiddlyWiki
Wiki link	[[Target|text]]	WikiWord	[[[Target|text]]]	[[text|Target]] WikiWord
External link	[URI text]	URI	[URI text]	[[text|URL]]
Image	[[File:W.png|…]]	URI	[[image W.png]]	[img[W.png]]
Table	{| …		…		|||| …		…	
…	…	…	…	…

EASY: Extract, Analyse, Synthesize

Source code ~~Source~~ Wiki text → Extract → Internal representation → Analyse → Internal representation → Synthesise → Results

P. Klint, T. van der Storm, J. Vinju. *EASY Meta-programming with Rascal*. GTTSE 2009, LNCS 6491, pp. 222–289. Springer, 2011.

Concrete wiki syntax migration

✓ EASY: Extract, Analyse, Synthesize.

✓ Extract = parse wiki pages

  ✓ write your own grammar

  ✓ use one of the existing grammars

✓ VVZ, *MediaWiki Grammar Recovery*, ACM CoRR, cs.MM, http://arxiv.org/abs/1107.4661, http://slps.sourceforge.net/zoo#wiki

Concrete wiki syntax example

module **MediaWiki**

syntax **WikiText** = **WikiSymbol***;
syntax **WikiSymbol** = **MWWord** w | **WikiLink** wl;
syntax **WikiLink**
  = simple: "[[" **PageName** destination "]]"
  | complex: "[[" **PageName** destination "|" **MWWord**+
description "]]" ;

lexical **MWWord** = ![\ \t\n\r\[\]]+ !>> ![\ \t\n\r\[\]] ;
lexical **PageName** = ![#\<\>\[\]|{}\t\n\r]+
                       !>> ![#\<\>\[\]|{}\t\n\r] ;
layout **MediaWiki** = [\ \t\n\r]* !>> [\ \t\n\r];

Concrete wiki syntax migration

✓ EASY: Extract, Analyse, Synthesize.

✓ Analyse

  ✓ correctness validation

  ✓ data mining and inference

  ✓ dealing with ontologies

✓ Synthesize

✓ pretty-print wiki text

Concrete wiki syntax example
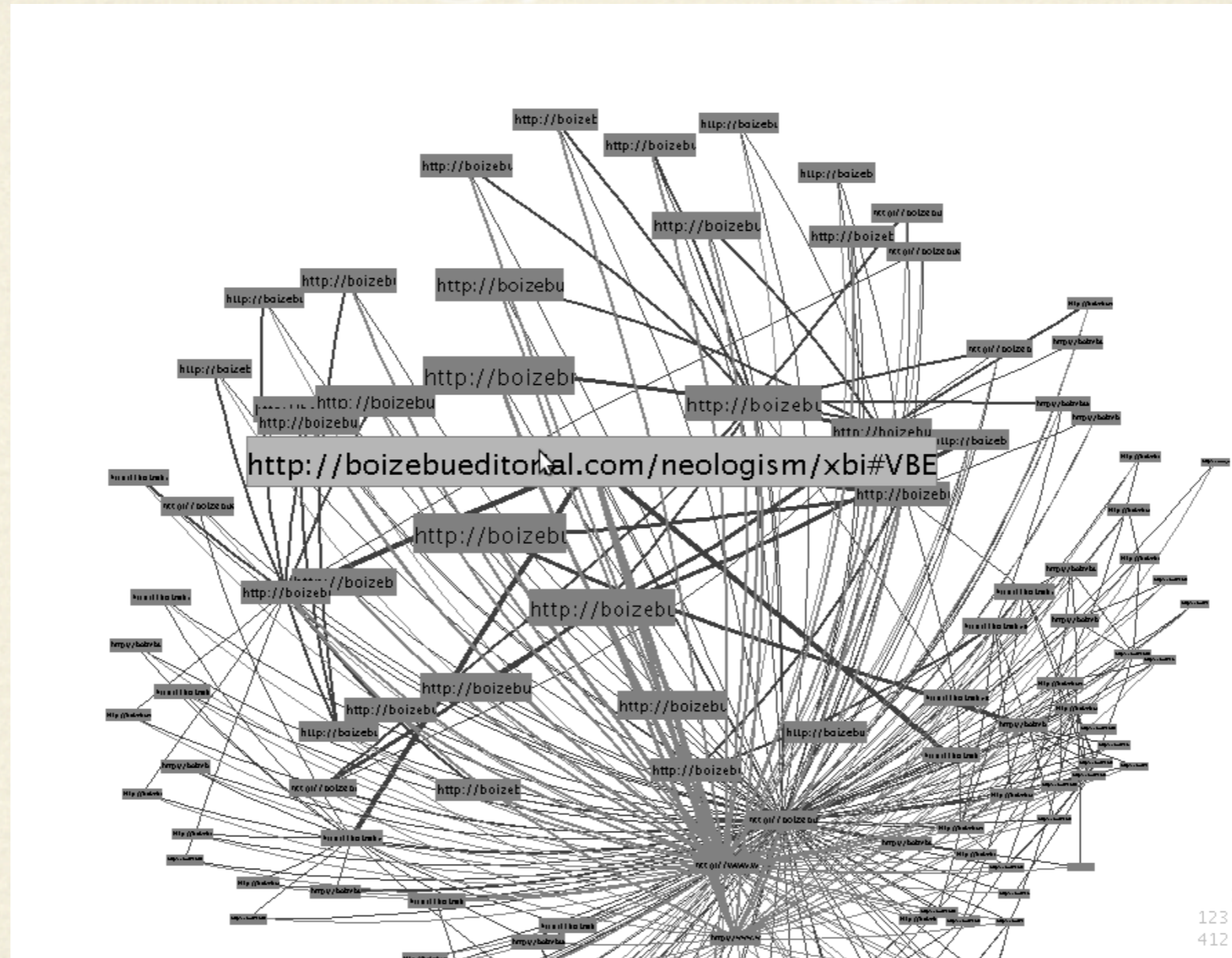
```
public str ppWT((WikiText)`<WikiSymbol* wss>`)
{return "<for(ws <- wss){><ppWS(ws)><}>"; }

public str ppWS((WikiSymbol)`<MWWord w>`)
{ return "<w>"; }

public str ppWS((WikiSymbol)`<WikiLink wl>`)
{
 if ( wl.prod.def.name == "simple" )
  return "[[<wl.destination>]]";
 else
  return "[[<wl.description>|<wl.destination>]]";
}
```
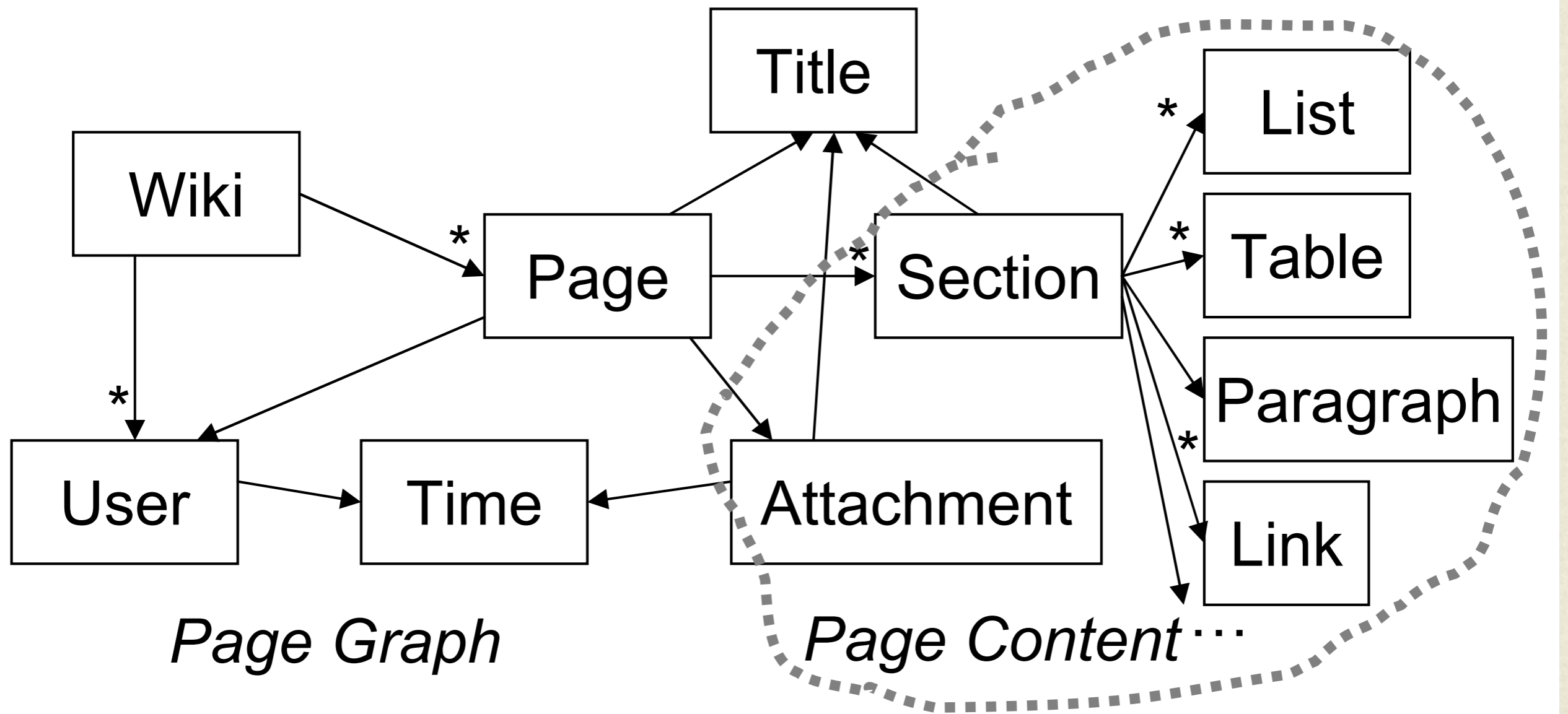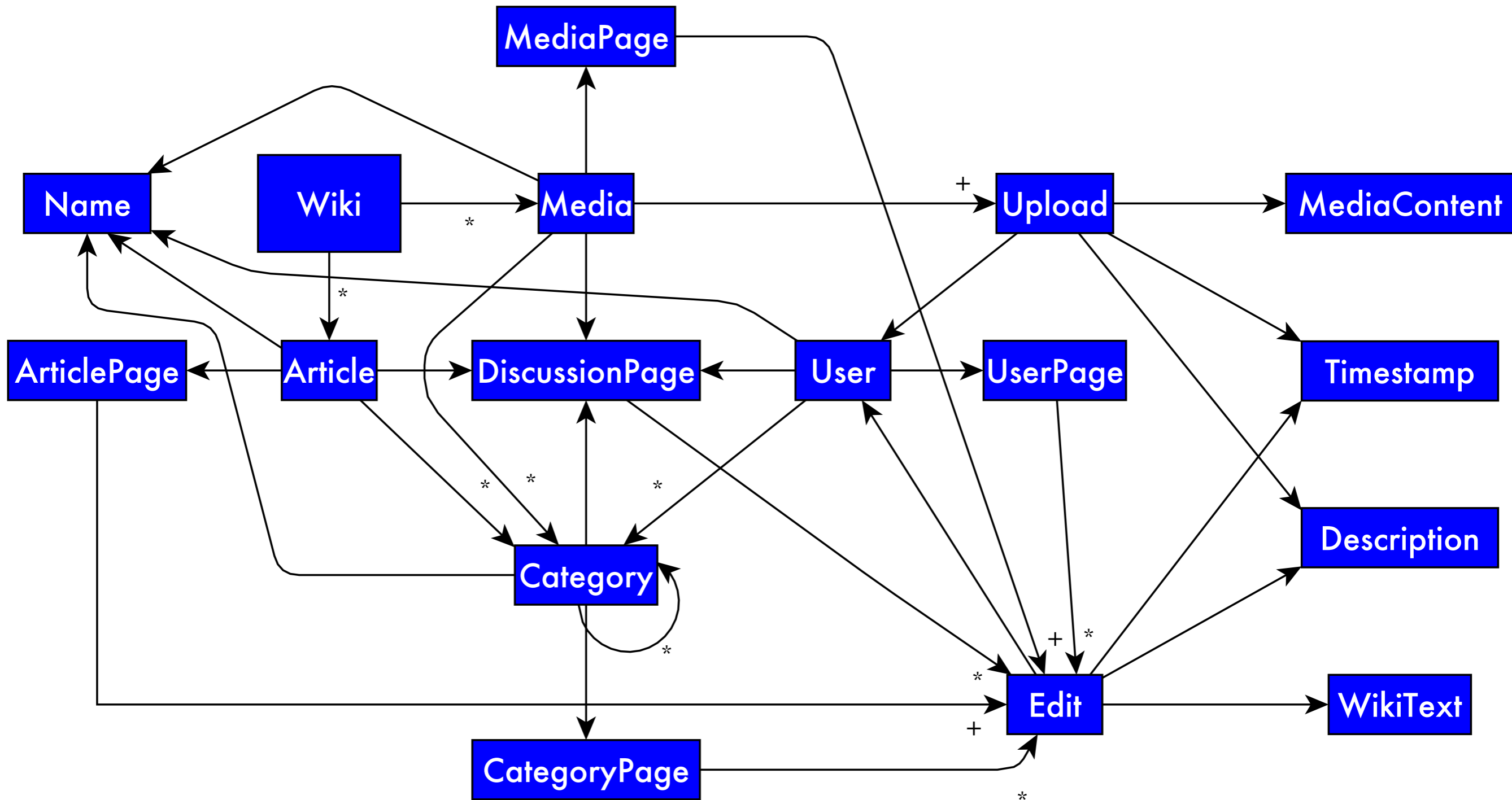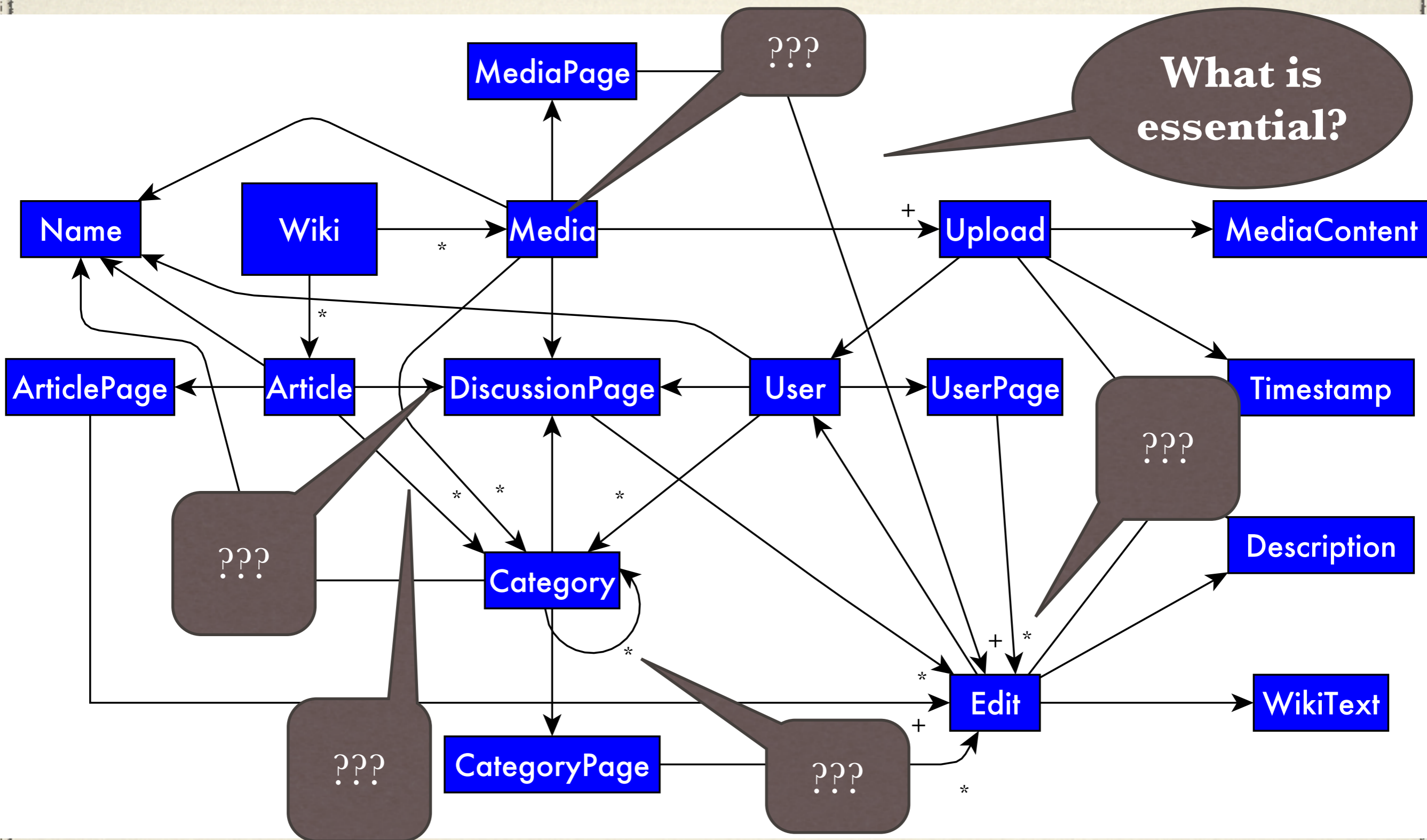
Ontology Migration

6 slides

Wiki ontology (Völkel&Oren)
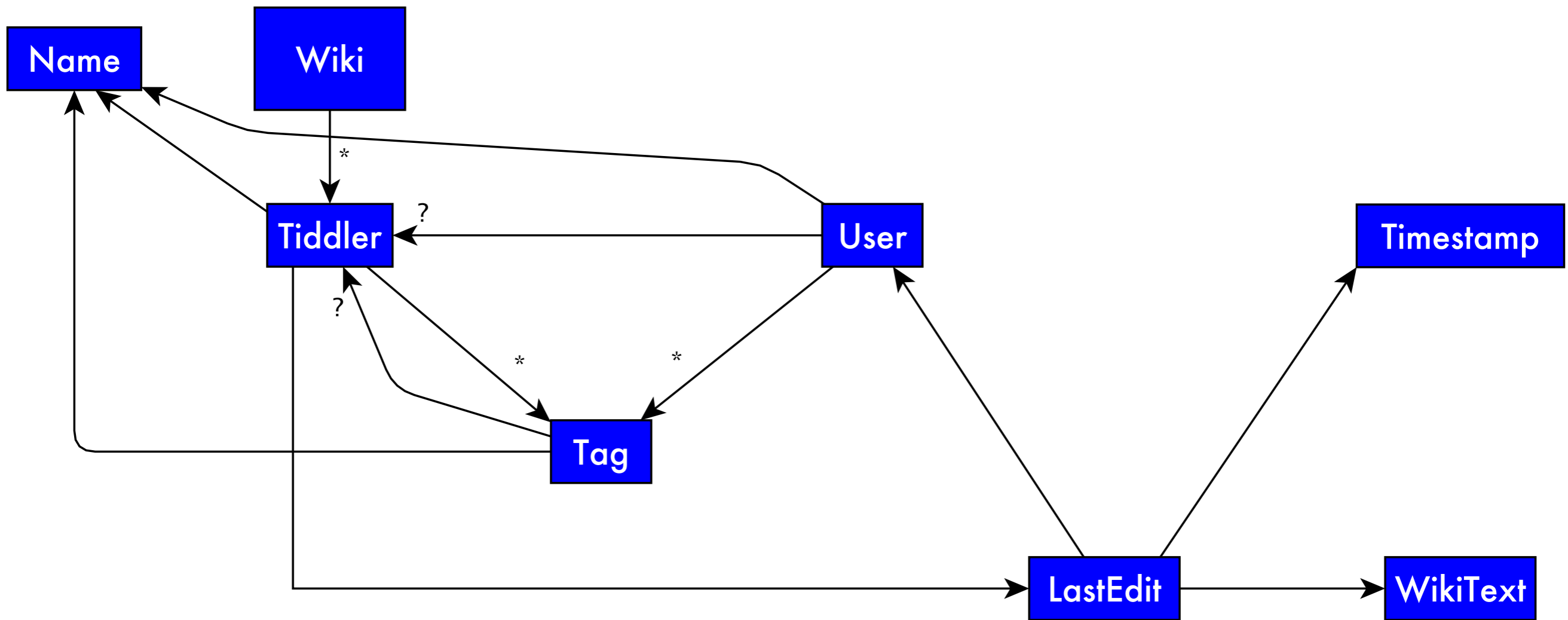

Table 1. A high-level view on the wiki data model

Wiki ontology: MediaWiki

Wiki ontology: MediaWiki

Wiki ontology: WikiDot

Wiki ontology: TiddlyWiki

Ontology mismatches

✓ Article: current content or history of revisions?

✓ Media: attachments, files, articles?

✓ Authorship: implicit, collaborative, explicit, none?

✓ Tagging: recursive, folksonomous, predefined?

✓ Illustration: framed singles, galleries, inlines, none?

✓ Level of detail: many small articles, several lengthy?

✓ User information: how easy to get? how personal?

Stilistical mismatches

✓ Domain specificity

   ✓ "in the fictional universe X, …"

✓ Wikification

   ✓ aggressive or minimalistic

✓ POV

   ✓ NPOV or expert opinion

✓ Proofreading or a lack thereof

✓ Notability, OR, …

Conclusion

✓ Prepare well for migration

✓ Inform the community

✓ Automate content migration

✓ Resolve ontology mismatches

✓ Minimise problems for wikipedians

✓ Encapsulate new content

✓ Proceed with the new wiki!

Thank you!

http://grammarware.net/

http://www.rascal-mpl.org/

Questions?