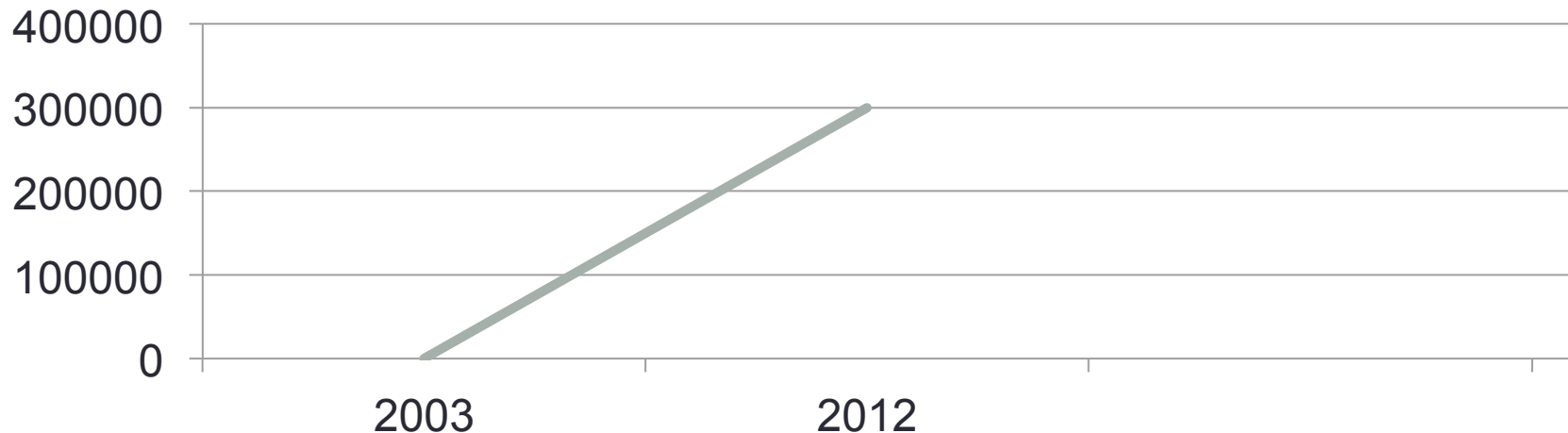


Lexical Computing Ltd
and the
Sketch Engine

Adam Kilgarriff

Figures

- Ten years old
- No borrowing or investors
- Turnover, GBP



- UK (LCL, me) and Cz (LCC, wholly owned)
 - development team of 7
 - also lecturers/students at Masaryk Uni, Brno

What we do: *corpora*

- Strap line:

Corpora for all

- Top corpus query tool
 - The Sketch Engine
- Web service
 - *also* The Sketch Engine
- Corpus building
 - To put in the Sketch Engine for all of our users
 - To customer specifications
- Corpus-derived lists:
 - Words, lemmas, word-tag pairs, n-grams, ...

Languages

- Sixty
- Major world languages
 - Multi-billion words
- Most others
 - 50 million and more
- Language-specific
 - Tokenisation
 - Lemmatisation
 - Map from text-word to dictionary-word, *invading* to *invade* (verb)
 - Part-of-speech tagging
 - *Collaboration with language experts*
 - Never trust a tool from someone who doesn't know the language

Customers

- Dictionary publishers
 - Oxford Univ Press, Cambridge Univ Press, Collins, Le Robert, Cornelsen, K Dictionaries, Shogakukan
- National Language Institutes
 - Bulgarian, Czech, Estonian, Irish, Netherlands, Slovak, Slovene
- Universities
 - Teaching, research, language teaching
 - Over 100 in many countries
- Individuals
- Tech companies
 - Mostly word lists
- Terminology/translation: WIPO

This meeting

- Interface
 - Between universities and vendors
 - Expertise, contacts (mainly academic)
 - corpora and tools for many languages
- Translation and terminology
 - Corpus modules to integrate with CAT tools
- Back-end data preparation

Thank you

adam@lexmasterclass.com

<http://www.sketchengine.co.uk>