

How well does research support KNP management objectives?

Barend Erasmus

School of Animal, Plant and
Environmental Sciences

University of the Witwatersrand

Overview

- Abstracts, authors and affiliations of all presentations and posters from 2003 to 2009
- Description of trends in numbers of authors and institutions
- Text mining
- Matching text mining outputs to the management objectives hierarchy

- Caveats with using abstracts as input data
 - Not all abstracts are equal!
 - Descriptive work vs research outputs
 - Not all research in SANPARKS (or just KNP) is presented

- Caveats with using the management objectives hierarchy as a guideline

- It is a management tool

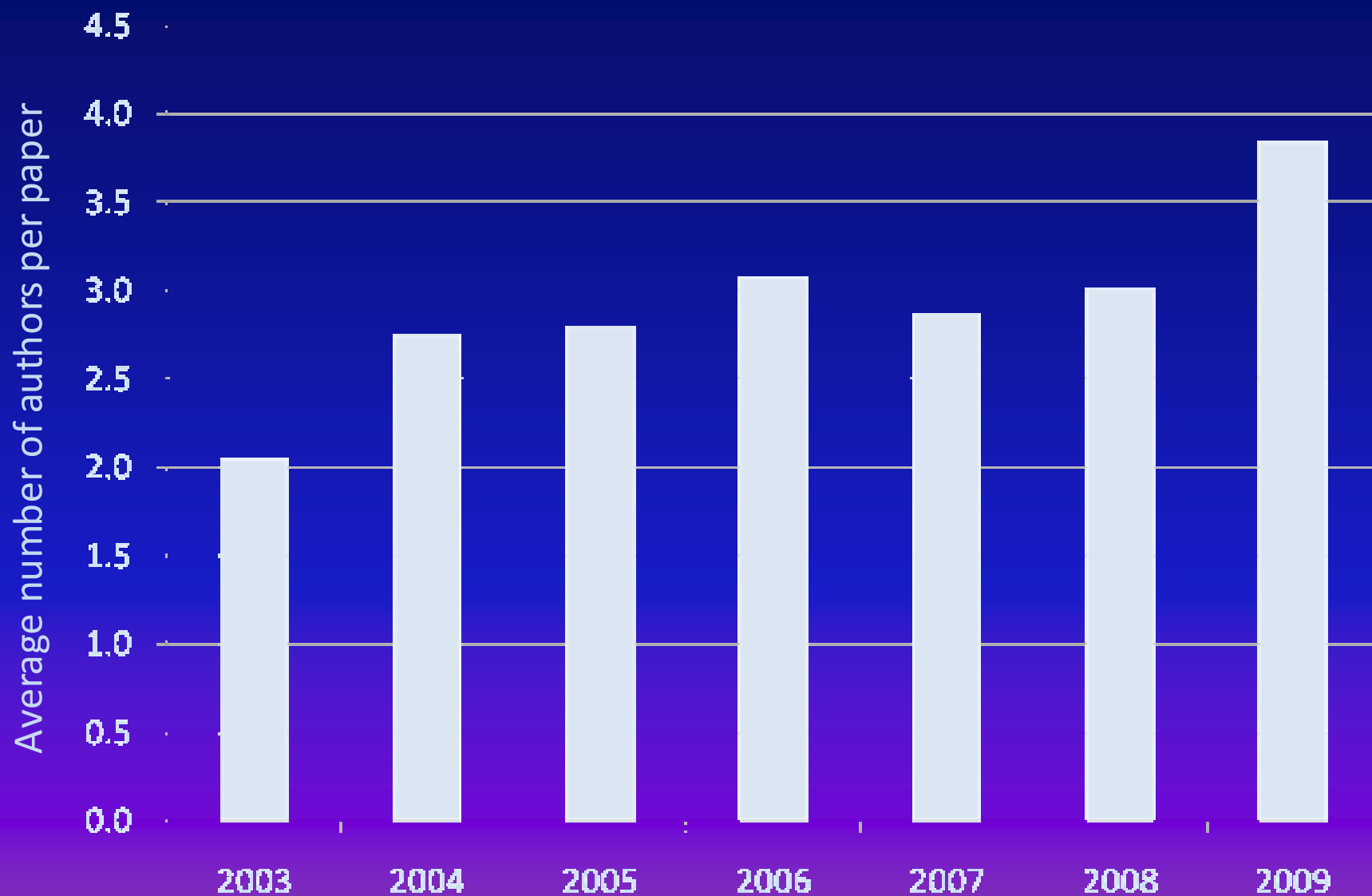
- Even though different objectives might occupy the same “branch” on the hierarchy, they may require very different resources to fulfil, i.e.

- “Develop understanding of silt-mediated nutrient aggregation and movement on floodplains”

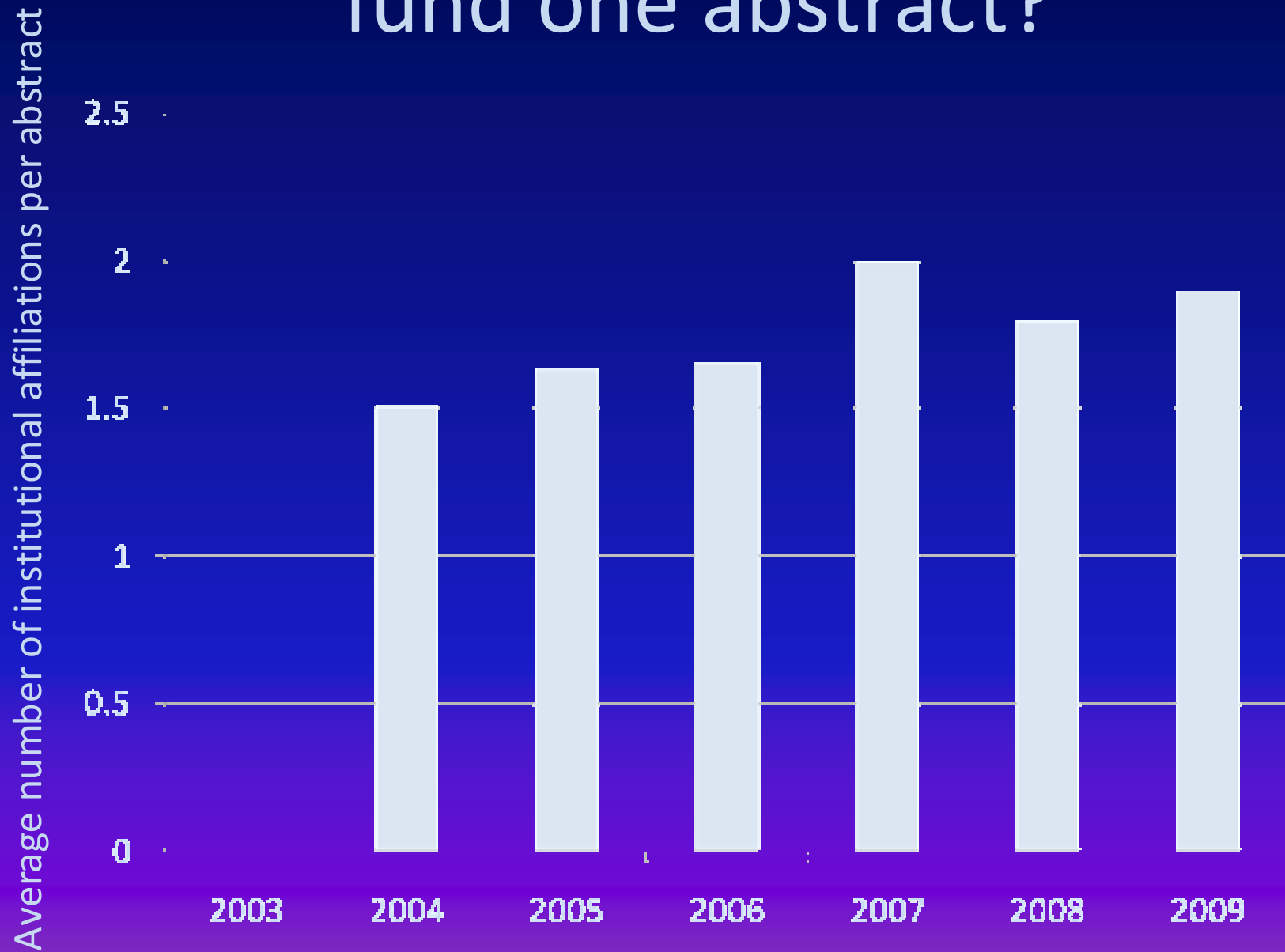
VS

- “Compile standardised, complete and updated taxa list”

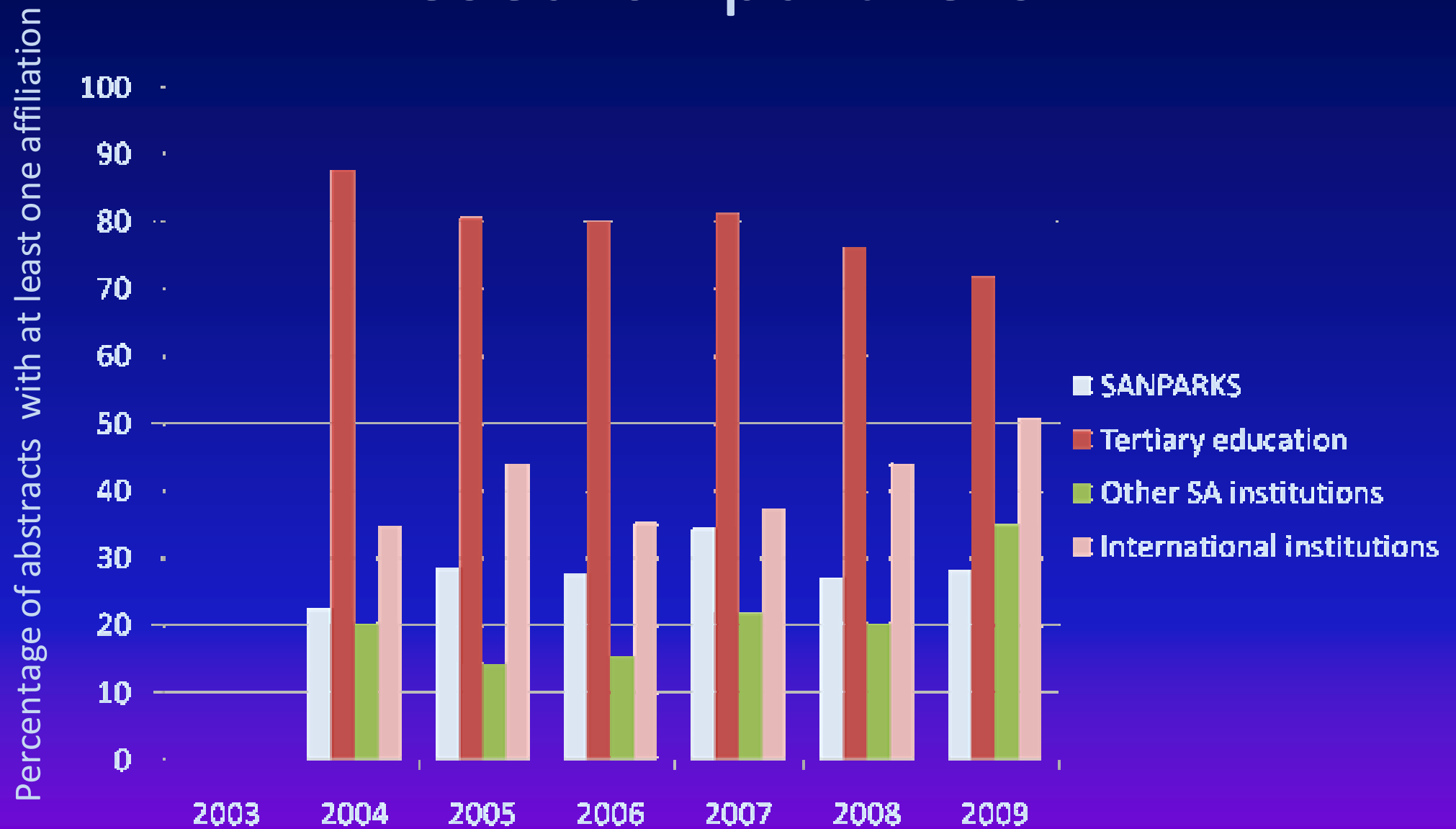
How many scientists does it take to write an abstract?



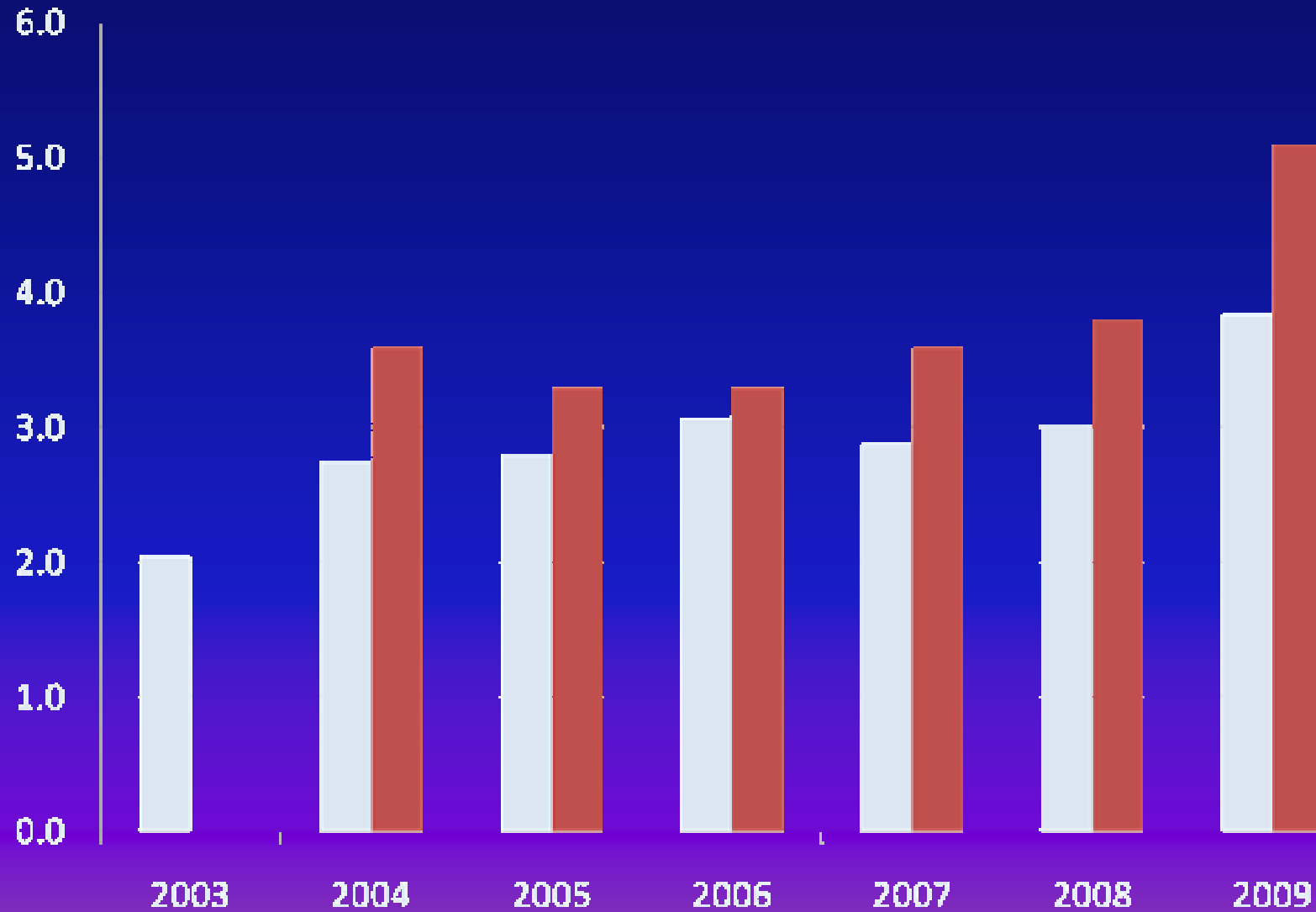
How many institutions does it take to fund one abstract?



Research partners



Average number of authors per paper, compared with average number of authors for international collaborations



Text mining

Purpose: To extract new meaning, generate new knowledge by using word associations in large text data bases.

- Extracting terms
- Counting terms
- Linking terms

But...

....how do you identify a term?

Text mining tool

- Frantzi, K., Ananiadou, S. and Mima, H. (2000) Automatic recognition of multi-word terms. *International Journal of Digital Libraries* 3(2), pp.117-132.
- TerMine service, to calculate the C value:
<http://www.nactem.ac.uk/software/termine/>

- Domain independent approach
- Linguistic and statistical analyses
 - Linguistic analysis lists candidate terms
 - Statistical analysis
 - the occurrence frequency of the candidate term
 - the frequency of the candidate term as part of other longer candidate terms
 - the number of these longer candidate terms
 - the length of the candidate term

Interpretation

- C values are comparable within years, not across years
- Term lists need an expert opinion:
 - Remove references to locations
 - For the purpose of matching with management objectives, remove very generic terms (ie “savanna ecosystem”)
 - Interpret the top 30 ranked terms (total number of terms 860 – 2000), and compare with objectives hierarchy

Example of output, for 2003 abstracts

Rank	Term	Score
1	kruger national park	29.114288
2	sabie river	10
2	burnt plot	10
4	south africa	9
5	woody vegetation	7.75
6	elephant impact	7
7	kruger park	6
7	riparian zone	6
7	savanna ecosystem	6
10	soil moisture	5.5
11	soil water	5
11	tree cover	5
11	nutrient cycling	5
11	annual burning	5
11	unburnt plot	5
16	bunch grass ecosystem	4.754888

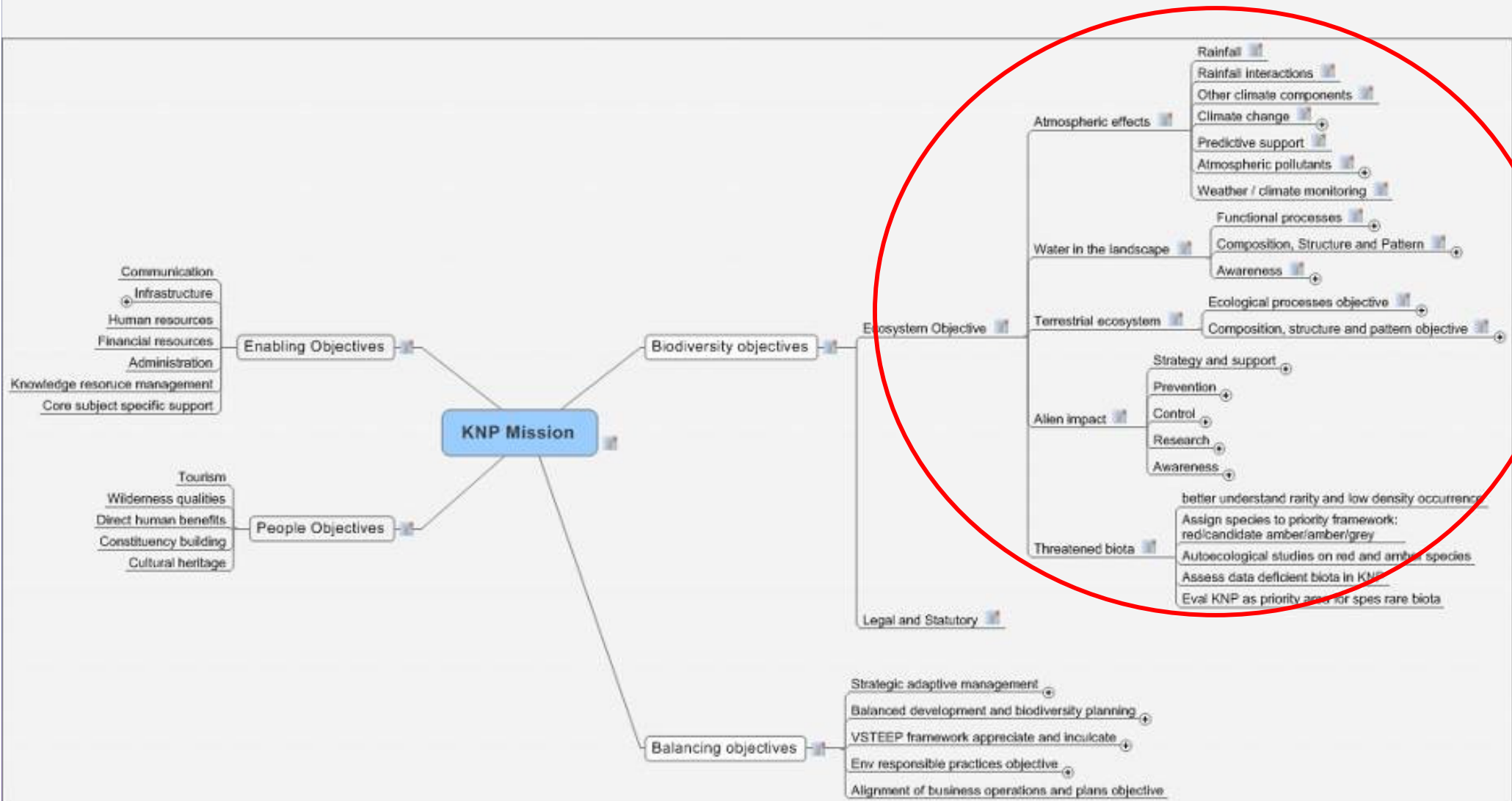
- Generic term recognition
- Preserve line breaks – distance between terms does matter

Top 30 ranked terms: some observations

- 2007 and 2008: the names of other parks start appearing (Gorongosa, Marakele, Tswalu, Limpopo, Mapungubwe)
- “Termite” appears only once in 2005, and “microbial biomass” only in 2003

TERM	First appearance in top 30
Adaptive management	2009
Scale	2006
Sodic site	2005
Heterogeneity	2004
Elephant	2003
Greenhouse gas	2009

The objectives management hierarchy



Objectives in blue have terms in the top 30 across all years – it is a research focus with lots of support over all analysis years

- Tree:grass ratios
- Determinants of woody cover
- Ecological role of fire
- Herbivore impacts on vegetation
- Impacts of water hole closure
- Plant-animal interactions

Ecosystem

Threatened biota

Autoecological studies on red and amber species

Eval KNP as priority area for spes rare biota

oring method

Objectives in pink have terms in the top 30 only in the first 3 or 4 years of the analysis period – it was a research focus that is now getting less attention

- The role of micro-organisms in KNP ecosystems
- Nutrient cycling: biotic effects
- Soil nutrient-mediated interactions between plants and animals
- Factors that affect decomposition rates, and its ecological importance
- Alien invasive plants (only one instance in top30)

Objectives in yellow have terms in the top 30 only in the 2 or 3 years of the analysis period – it is a developing research focus

- The role of indigenous disease to maintain savanna dynamics, ie BTB
- Factors that affect prey selection
- Reciprocal impacts between predator and prey

Take home messages - methods

- Text mining yield plausible results but a more objective ranking cut-off selection will enable matching of terms and objectives lower down the hierarchy
- There is still a lot of information in the lower ranked terms, especially to identify emerging topics.
- Abstracts is probably not the best input data for this analysis:
 - Not representative of all research and management activities that contribute to fulfilling the needs of the objectives hierarchy.
 - No peer review, so less rigorously written, especially wrt reporting results

Take home message – research links

- Text mining identified trends that are consistent with known developments in research foci
- There is considerable overlap between the fields of research being done in KNP, and management needs
- In most abstracts, management implications of research findings are only implied.
- It is not clear from this data whether even the research projects with definitive management outcomes were completed on management-relevant time- and spatial scales
- ...it is unlikely, since there is no mechanism by which South African research funding cycles will match KNP management needs.