

The Present & Future of Google Books

Dr. James Crawford

**Engineering Director for the
Google Books Project**

**Google, Inc.
Mountainview, CA**

Keynote

ETD 2010

Austin, TX

June 17, 2010

**Dr. Crawford was
Executive Vice president at
Composite Software,
an enterprise software startup
in the Bay area.**

**During his four years at Composite,
his teams delivered next generation
data integration technology to some of
the world's largest financial institutions,
manufacturers, and
multiple government agencies.**

**Before joining Composite Software,
Dr. Crawford spent three years at
NASA's Ames Research Center as
lead for autonomy and robotics.
One of his teams delivered the
optimized activity
planner used by both the
Opportunity and Spirit Mars rovers.**

**Prior to joining NASA, he was the
optimization architect for the
I2 Technologies
supply chain planner,
co-founded the
Computational Intelligence Research
Laboratory (CIRL)
at the University of Oregon, and
worked at
AT&T Bell Laboratories.**

The Google Books project

has the modest goal of scanning all of the world's books, converting them to digital form, and making them searchable as easily as web pages.

To date over twelve million books containing over four billion pages have been scanned and digitized.

The challenges have been one part technical, including OCR for dozens of languages, and one part legal (copyright laws were inherited from a different age).

As the corpus emerges, and the legal challenges are overcome, books are becoming more accessible than ever before and new opportunities are arising in areas ranging from social collaboration to linguistic analysis.

Relevance for ETD 2010:

- 1) Scanning backfiles of ETDs**
- 2) Gail Clement hypothesis of 1939-1976 ETDs in public domain.**
- 3) Current ETDs into Google Scholar:
Recent communication between
Thom Hickey, OCLC, and
Anurag Acharya, Google Scholar:
Writing parsers for
ETD-MS and MARC
Crawling using our metadata, provided from
NDLTD Union Catalog**