Seamless access to the world’s open access research papers

Nancy Pontika

core.ac.uk
In this talk

• What is CORE?
• Who is CORE for?
• What services does CORE offer?
Mission

Aggregate all open access articles across relevant data sources worldwide, enrich this content and provide seamless access to it through a set of data services …
World’s largest open access articles collection

> 135,116,215 Metadata
> 105,290,852 Abstracts
> 11,817,841 Hosted full texts
> 78,781,218 Links to full text

60TB of data
3,788 Data providers
~ 50k Journals

CORE API
CORE Dataset
CORE Search
CORE Repository Dashboard
CORE Publisher Connector
CORE SDKs
CORE Recommender
CORE Analytics Dashboard
Global reach
Harvesting

1. OAI-PMH
2. ResourceSync
3. Bespoke APIs of Publishers
4. Custom discovery service
CORE Pipeline

- Metadata download, extraction and harmonisation
- Full text download
- Text extractions and sections extraction
- Metadata validation
- Thumbnails generation
- References and citation contexts extraction
- API enrichment (e.g. finding DOIs, linking to other systems)
- Document type classification
- Deduplication
- Indexing
- Exposing (data dumps, API, ResourceSync)
CORE Search

Aggregating the world’s open access research papers

We offer seamless access to millions of open access research papers, enrich the collected data for text-mining and provide unique services to the research community.

Read more  
Join us
Energy efficient data mining scheme for high dimensional data

By Mohammad Alwadi and Girija Chetty
Repository: University of Canberra Research Repository | 2015

Smart phone based data mining for human activity recognition

By Girija Chetty, Matthew White and Farnaz Akther
Repository: University of Canberra Research Repository | 2015

Analysing qualitative data from virtual worlds: using images and text mining

By Brad James Austin McKenna, Michael Myers and Lesley…
CORE Search

- Faceted searching
- What you find is what you can download
- We really host the full text content
- We are not here to replace Google Scholar, we serve a different community
- >22 million yearly unique visitors
CORE Recommender

• Recommending relevant content to users from across the CORE’s content
• Recommender plugin for repositories
• 23 repositories using it, e.g. Cambridge University
CORE Recommender

- Recommending relevant content to users from across the CORE’s content
- Recommender plugin for repositories
- 23 repositories using it, e.g. Cambridge University
CORE API, data dumps, SDKs

- Enable others to programmatically interact with CORE data.
- Essential for developing applications that make use of information about research papers
- Supporting a variety of use cases
- >500 registered users, including companies for CORE API
- Use:
  - CORE API to interact with CORE in real time
  - CORE Data Dumps to run batch process, develop prototype applications and do research
  - CORE SDKs to gain easy access to CORE API from several programming languages
  - CORE FastSync
CORE API, data dumps, SDKs

• It is beyond human capacities to read all scientific literature
• Example use cases in which CORE is applied:
  • Improving discovery
  • Plagiarism detection
  • Question answering in science
  • Literature based discovery
  • Fact checking and detection of misinformation
  • Analysing research trends
  • Finding experts in a particular domain
  • Research evaluation and scientometrics
  • Exploratory and visual search
  • ...
CORE Repository Dashboard

- Helps repositories to:
  - Manage content aggregated by CORE
  - Identify and resolve issues in the configuration of their repository
- >60 registered repositories
CORE Analytics Dashboard

• A tool for analytical evaluation of universities research outputs based primarily on CORE data.
• Research Impact insights on institutional level, comparison with other institutions or group of institutions
• Based on citation data, social impact data
• Data sources:
  – CORE dataset (primary pivot)
  – MAG dataset
  – Crossref Event Data
  – Mendeley
CORE Analytics Dashboard

- Helps you analyse the impact of research at your institution.
- See how you perform against competitors.
- Slice and dice by discipline, field, any kind of a query.
CORE Analytics Dashboard: sneak peek

- Customizable area
- Pre-defined widgets of standard data reports
Key customers/partners

- Main source of data for **Naver Academic**.
- **IRIS.AI**: A start up of 20 people entirely developing tools to help researchers be more productive. Highly dependent on CORE API.
- **ProQuest**: CORE supplies data about OA publications to improve discovery of OA content in Primo and Summon (new) additional discovery collaborations on the way.
- Providing a key API for **OA Button**.
Our contribution to the community (examples)

- Harvesting and delivering data from UK repositories supporting RIOXX to OpenAIRE (every 2 weeks) on behalf of Jisc.
- **OpenAIRE** to use CORE Publisher Connector.
- Delivering repository statistics data (including historical information) to OpenDOAR (new).
- Providing statistics on the growth of OA and disciplinary trends (in particular to British organisations).
- **IRUS-UK**: IRUS has created in collaboration with CORE the IRUS-CORE service which integrates the capturing of download data from repositories with CORE.
- **OpenMinTeD**: EU eInfrastructure project for TDM of research data. CORE the largest supplier of data.
- Driving innovation in this area (**EOSC/ResourceSync**), NGR.
Conclusions

• Global world’s largest aggregator of open access papers
• A suite of data services for a variety of stakeholders
• Large user base
• Moving fast from research to a product. Always innovate.
Thank you!

CORE
Bringing science to all

@oacore
theteam@core.ac.uk
core.ac.uk/team