
dd_api Documentation

Release 1.0

Yamuna Krishnamurthy

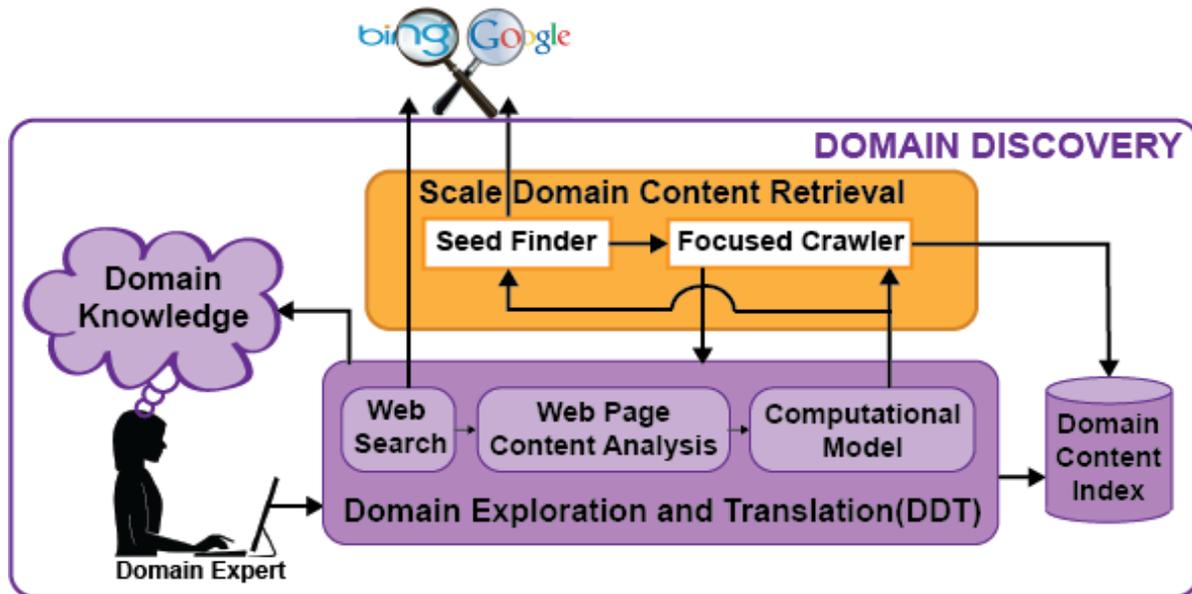
Jun 11, 2019

Contents

1	Contents	3
1.1	Installation	3
1.2	DD API Operations	4
2	Links	5
3	Indices and tables	7

Domain Discovery is the process of acquiring, understanding and exploring data for a specific domain. Some example domains include human trafficking, illegal sale of weapons and micro-cap fraud. While acquiring knowledge about a domain humans usually start with a conception of that domain. This conception is based on prior knowledge of parts of the domain. The process of gaining a more complete knowledge of the domain involves using this prior knowledge to obtain content that provides additional information about that domain that was previously unknown. This new knowledge of the domain now becomes prior knowledge leading to an iterative process of domain discovery as illustrated in Figure 2. The goals of this iterative domain discovery process are:

- complete the human’s knowledge of the domain
- acquire sufficient content that captures the human cognition of the domain to translate into a computational model



The Domain Discovery API formalizes the human domain discovery process by defining a set of operations that capture the essential tasks that lead to domain discovery on the Web as we have discovered in interacting with the Subject Matter Experts (SME)s. The API facilitates:

- Creation of different user interfaces to satisfy different DD needs
- Configure and understand different DD workflows
- Scripting DD

1.1 Installation

Building and deploying the Domain Discovery can be done using its Makefile to create a local development environment. The conda build environment is currently only supported on 64-bit OS X and Linux.

First install conda, either through the Anaconda or miniconda installers provided by Continuum. You will also need Git, a Java Development Kit and Maven. These are system tools that are generally not provided by conda.

Clone the DD API repository and enter it:

```
>>> git clone https://github.com/ViDA-NYU/domain_discovery_API
>>> cd domain_discovery_API
```

Use the *make* command to build DD API and download/install its dependencies.

```
>>> make
```

Now you can use the API

1.2 DD API Operations

1.2.1 Acquire Content

1.2.2 Annotate Content

1.2.3 Summarize Content

1.2.4 Organize Content

1.2.5 Filter Content

1.2.6 Generate Model

CHAPTER 2

Links

- [GitHub repository](#)

CHAPTER 3

Indices and tables

- `genindex`
- `modindex`
- `search`