

Routing Policies

Typically, when a worker asks for a batch, we will supply a single batch (if we have one). But we can also configure more workers with different "tags" meaning that the particular worker needs just "tagged" batches. For example one worker can ask only for batches tagged with the region "EU" and the other one for batches tagged with the region "US".

The worker configuration parameter:

- tags - the list of worker tags (for example "EU,US")

On the manager side we define first Routing policies and we can assign ids of those policies to:

- seeds
- connections

When a manager prepares batches for workers it collects tags from routing policies assigned to crawled seeds and connections and assign them to the batches prepared for sending.. The decision making by the manager when the worker asks for batches goes like this:

- If the batch has no tags, it can go anywhere
- Otherwise do a simple check - if the number of tags given by a worker is less than the number in the batch, it can't match them all. Therefore the batch is not suitable for the worker
- Otherwise, every batch tag must occur in the worker tag