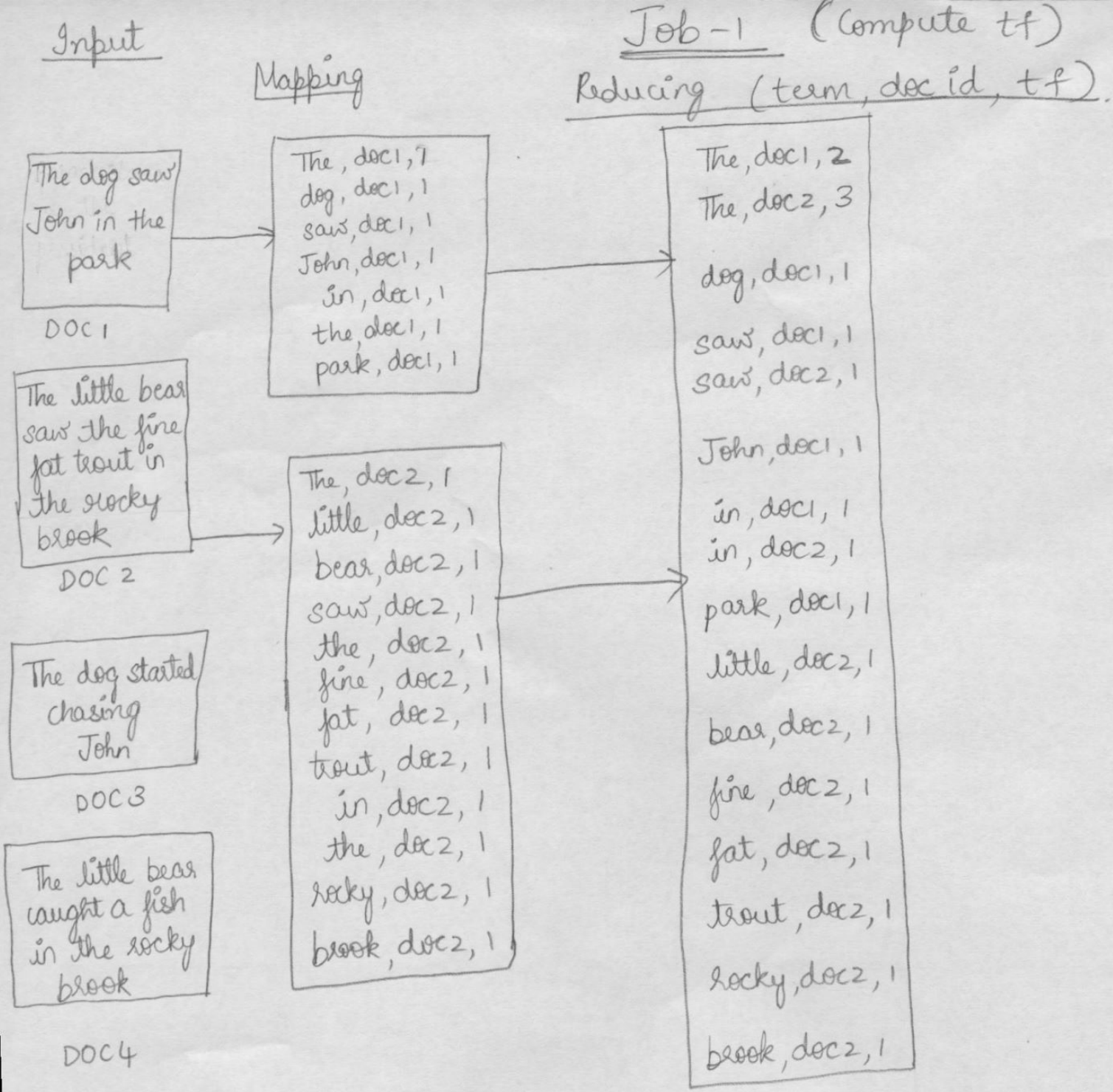
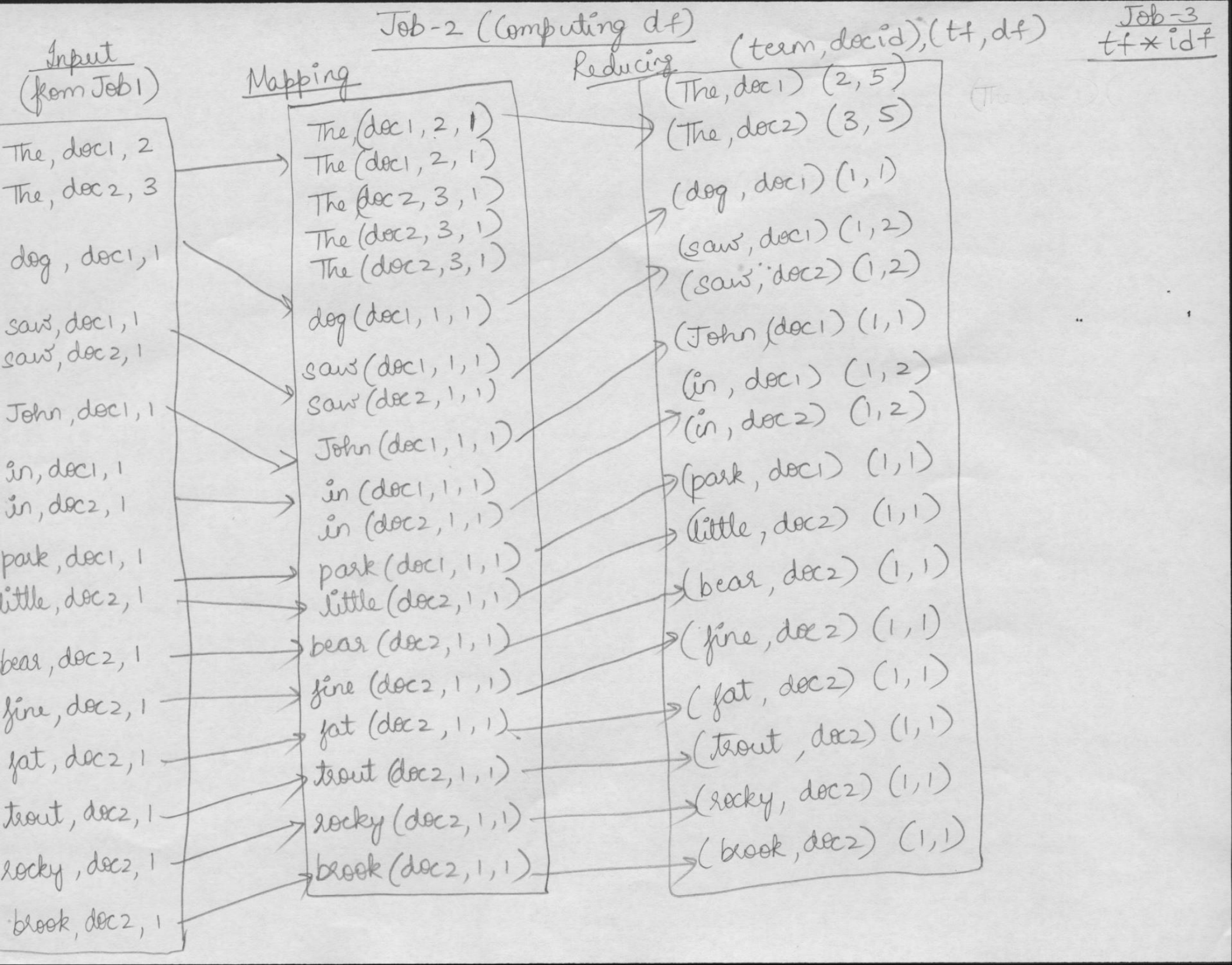

Input
Mapping
Job-1 (Compute tf)
Reducing (term, doc id, tf).
The dog saw John in the park
DOC 1
The little bear saw the fine fat trout in the rocky brook
DOC 2
The dog started chasing John
DOC 3
The little bear caught a fish in the rocky brook
DOC 4
The, doc1, 1
dog, doc1, 1
saw, doc1, 1
John, doc1, 1
in, doc1, 1
the, doc1, 1
park, doc1, 1
The, doc2, 1
little, doc2, 1
bear, doc2, 1
saw, doc2, 1
the, doc2, 1
fine, doc2, 1
fat, doc2, 1
trout, doc2, 1
in, doc2, 1
the, doc2, 1
rocky, doc2, 1
brook, doc2, 1
The, doc1, 2
The, doc2, 3
dog, doc1, 1
saw, doc1, 1
saw, doc2, 1
John, doc1, 1
in, doc1, 1
in, doc2, 1
park, doc1, 1
little, doc2, 1
bear, doc2, 1
fine, doc2, 1
fat, doc2, 1
trout, doc2, 1
rocky, doc2, 1
brook, doc2, 1


Job-2 (Computing df)
Input (from Job1)
Mapping
Reducing
(term, docid), (tf, df)
Job-3 tf * idf
The, doc1, 2
The, doc2, 3
dog, doc1, 1
saw, doc1, 1
saw, doc2, 1
John, doc1, 1
in, doc1, 1
in, doc2, 1
park, doc1, 1
little, doc2, 1
bear, doc2, 1
fine, doc2, 1
fat, doc2, 1
trout, doc2, 1
rocky, doc2, 1
brook, doc2, 1
The, (doc1, 2, 1)
The (doc1, 2, 1)
The (doc2, 3, 1)
The (doc2, 3, 1)
The (doc2, 3, 1)
dog (doc1, 1, 1)
saw (doc1, 1, 1)
saw (doc2, 1, 1)
John (doc1, 1, 1)
in (doc1, 1, 1)
in (doc2, 1, 1)
park (doc1, 1, 1)
little (doc2, 1, 1)
bear (doc2, 1, 1)
fine (doc2, 1, 1)
fat (doc2, 1, 1)
trout (doc2, 1, 1)
rocky (doc2, 1, 1)
brook (doc2, 1, 1)
(The, doc1) (2, 5)
(The, doc2) (3, 5)
(dog, doc1) (1, 1)
(saw, doc1) (1, 2)
(saw, doc2) (1, 2)
(John, doc1) (1, 1)
(in, doc1) (1, 2)
(in, doc2) (1, 2)
(park, doc1) (1, 1)
(little, doc2) (1, 1)
(bear, doc2) (1, 1)
(fine, doc2) (1, 1)
(fat, doc2) (1, 1)
(trout, doc2) (1, 1)
(rocky, doc2) (1, 1)
(brook, doc2) (1, 1)

Job-3 Compute (tf * idf)

$tf * \log\left(\frac{N}{df}\right)$ (term, docid, tf*idf)

(The, doc1) (2, 5)
(The, doc2) (3,5) → The, doc1, $2 * \log\left(\frac{2}{2}\right)$ → (The, doc1, doc2, 0)

(dog, doc1) (1, 1) → dog, doc1, $1 * \log\left(\frac{2}{1}\right)$ → (dog, doc1, 0.301)

(saw, doc1) (1, 2)
(saw, doc2) (1,2) → saw, doc1, doc2, $2 * \log\left(\frac{2}{2}\right)$ → (saw, doc1, doc2, 0)

(John, doc1) (1, 1) → John, doc1, $2 * \log\left(\frac{2}{1}\right)$ → (John, doc1, doc2, 0.301)

(in doc1) (1, 2)
(in doc2) (1,2) → in, doc1, doc2, $2 * \log\left(\frac{2}{2}\right)$ → (in, doc1, doc2, 0.301)

(park, doc1) (1, 1) → park, doc1, $2 * \log\left(\frac{2}{1}\right)$ → (park, doc1, 0.301)

(little, doc2) (1, 1) → little, doc2, $2 * \log\left(\frac{2}{1}\right)$ → (little, doc2, 0.301)

(bear, doc2) (1,1) → bear, doc2, $2 * \log\left(\frac{2}{1}\right)$ → (bear, doc2, 0.301)

(fine, doc2) (1,1) → fine, doc2, $2 * \log\left(\frac{2}{1}\right)$ → (fine, doc2, 0.301)

(fat, doc2) (1,1) → fat, doc2, $2 * \log\left(\frac{2}{1}\right)$ → (fat, doc2, 0.301)