# Generelle Semantik

Jens Dittrich
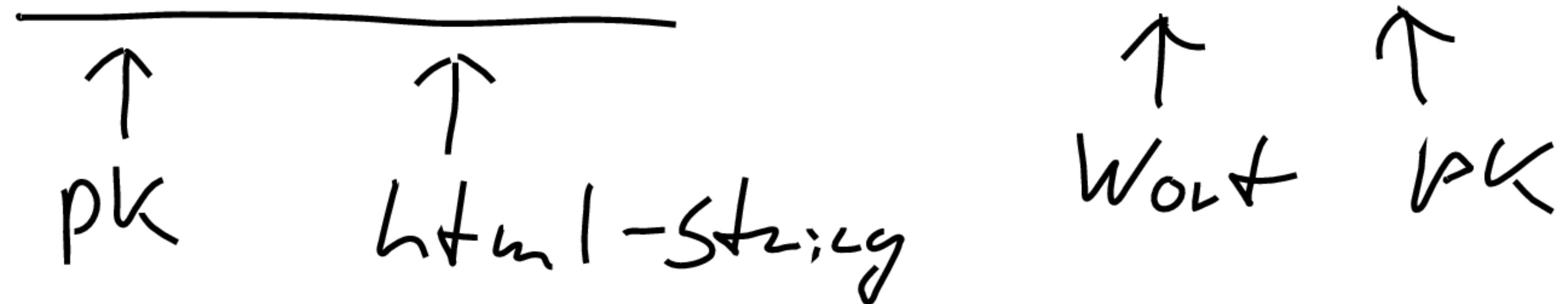
42

map(record) -> set of (ikey, ivalue)

reduce(ikey, set of ivalue) -> output

42, 3

42, 4

42, 5

42, {3,4,5}

# Beispiel: Web-Index Erstellung

map( (docID, document) ) -> set of (term, docID)

pk html-String Wort pk

reduce(term, set of docID) -> (term, (posting list of docID, count))

```
map( (44,
      "This is text on a website!")
)
->
{
        (`This´, 44),
        (`is´, 44),
        (`text´, 44),
        (`on´, 44),
        (`a´, 44),
        (`website´, 44)
}
```

```
map( (42,
       ˝This is just another website!˝)
)
->
{
        (`This˝, 42),
        (`is˝, 42),
        (`just˝, 42),
        (`another˝, 42),
        (`website˝, 42)
}
```

```
map( (43,
      ˝One more boring website!˝)
)
->
{

        (`One˝, 43),
        (`more˝, 43),
        (`boring˝, 43),
        (`website˝, 43)

}
```

reduce(``This´´,
        {42,
          43}
)
->
(``This´´, ([42, 43], 2))

```
reduce(`is´,
        {42,
         43}
)
->
(`is´, ([42, 43], 2))
```

reduce(`boring´,
        {43}
)
->
(`boring´, ([43], 1))
_____

etc.

# Beispiel: Gruppierung mit Aggregation

SELECT      erfahrung, count(*), max(gehalt)

FROM      mitarbeiter

WHERE      gehalt<50000

GROUP BY    erfahrung

map() und reduce()

Mitzeberibe

map(record) -> set of (ikey, ivalue):

$$\text{if } (gehalt < 50000)$$

$$\text{emit } (\underline{erfahrung}, \underline{gehalt});$$

$$\phantom{\text{emit } (}ikey \qquad\quad ivalue$$

reduce(ikey, set of ivalue) -> output:

$A =$

```
int maxGehalt = -42;
for each value in ivalue:
    maxGehalt = max(maxGehalt, value);
emit (ikey, set.size(), maxGehalt);
```

# Beispiel: HAVING

SELECT      erfahrung, count(*), max(gehalt)

FROM        mitarbeiter

WHERE       gehalt<50000

GROUP BY    erfahrung

HAVING      count(*)>2

map() und reduce()

map(record) -> set of (ikey, ivalue):

wie oben

reduce(ikey, set of ivalue) -> output:

if ( set . size () > 2 )

A