SAS Hash-哈希

Hash，一般翻译做散列、杂凑，或音译为哈希，是把任意长度的输入（又叫做预映射pre-image）通过散列算法变换成固定长度的输出，该输出就是散列值。

这种转换是一种压缩映射，也就是，散列值的空间通常远小于输入的空间，不同的输入可能会散列成相同的输出，所以不可能从散列值来确定唯一的输入值。

简单的说就是一种将任意长度的消息压缩到某一固定长度的消息摘要的函数。

In [1]:
```
/* DETAIL */
DATA DETAIL;
INPUT
    DATE DATE9.
    AREACODE
    PHONENUM $8.
    TOACRONYM $
    FROMACRONYM $
    SECRETCODE
;
FORMAT
    DATE YYMMDD10.
;
DATALINES;
21DEC2006 407 312-9088 AFAIK DQMOT 103
21DEC2006 407 324-6674 BEG TU 101
21DEC2006 407 312-9088 BFN SYS 101
21DEC2006 407 312-9088 BTDT IHU 102
22DEC2006 407 312-9088 C&G AFAIK 103
;

PROC PRINT DATA = DETAIL;
RUN;
```

SAS Connection established. Subprocess id is 3075

Out[1]:

The SAS System

Obs	DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE
1	2006-12-21	407	312-9088	AFAIK	DQMOT	103
2	2006-12-21	407	324-6674	BEG	TU	101
3	2006-12-21	407	312-9088	BFN	SYS	101
4	2006-12-21	407	312-9088	BTDT	IHU	102
5	2006-12-22	407	312-9088	C&G	AFAIK	103

```
In [2]:  /* LOOKUP_1 */
         DATA LOOKUP_1;
         INPUT
             ACRONYM $
             MEANING $30.
         ;
         DATALINES;
         AFAIK AS FAR AS I KNOW
         AFK AWAY FROM KEYBOARD
         ASAP AS SOON AS POSSIBLE
         BEG BIG EVIL GRIN
         BFN BYE FOR NOW
         BTDT BEEN THERE, DONE THAT
         DQMOT DON'T QUOTE ME ON THIS.
         IHU I HATE YOU
         SYS SEE YOU SOON
         ;

         PROC PRINT DATA = LOOKUP_1;
         RUN;
```

Out[2]:

The SAS System

Obs	ACRONYM	MEANING
1	AFAIK	AS FAR AS I KNOW
2	AFK	AWAY FROM KEYBOARD
3	ASAP	AS SOON AS POSSIBLE
4	BEG	BIG EVIL GRIN
5	BFN	BYE FOR NOW
6	BTDT	BEEN THERE, DONE THAT
7	DQMOT	DON'T QUOTE ME ON THIS.
8	IHU	I HATE YOU
9	SYS	SEE YOU SOON

RESULT:

DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
21DEC2006	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
21DEC2006	407	324-6674	BEG	TU	101	BIG EVIL GRIN
21DEC2006	407	312-9088	BFN	SYS	101	BYE FOR NOW
21DEC2006	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
22DEC2006	407	312-9088	C&G	AFAIK	103	

```
In [3]:  /*
             ——F1——
             MERGE
         */
         PROC SORT
             DATA = DETAIL
             OUT = _T1
         NODUPKEY;
         BY TOACRONYM;
         RUN;

         PROC SORT
             DATA = LOOKUP_1
             OUT = _T2(
                 RENAME = (
                     ACRONYM = TOACRONYM
                     MEANING = TOMEANING
                 )
             )
         NODUPKEY;
         BY ACRONYM;
         RUN;

         DATA T;
         MERGE _T1(IN=A) _T2(IN=B);
         BY TOACRONYM;
         IF A;
         RUN;

         PROC DELETE DATA = _T1 _T2;
         RUN;

         PROC PRINT DATA = T;
         RUN;
```

Out[3]:

The SAS System

Obs	DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
1	2006-12-21	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
2	2006-12-21	407	324-6674	BEG	TU	101	BIG EVIL GRIN
3	2006-12-21	407	312-9088	BFN	SYS	101	BYE FOR NOW
4	2006-12-21	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
5	2006-12-22	407	312-9088	C&G	AFAIK	103	

```
In [4]: /*
            --F2--
            SQL
        */
        PROC SQL;
        SELECT
            A.*,
            B.MEANING AS TOMEANING
        FROM DETAIL AS A LEFT JOIN LOOKUP_1 AS B
        ON A.TOACRONYM = B.ACRONYM
        ;
        QUIT;
```

Out[4]:

The SAS System

DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
2006-12-21	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
2006-12-21	407	324-6674	BEG	TU	101	BIG EVIL GRIN
2006-12-21	407	312-9088	BFN	SYS	101	BYE FOR NOW
2006-12-21	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
2006-12-22	407	312-9088	C&G	AFAIK	103	

```
In [5]: /*
            --F3--
            HASH (RIGHT)
        */
        DATA RESULTS;
        IF 0 THEN SET LOOKUP_1;
        DROP ACRONYM MEANING RC;
        IF _N_ = 1 THEN DO;
            DECLARE HASH HASHLOOKUP(DATASET:'LOOKUP_1');
            HASHLOOKUP.DEFINEKEY('ACRONYM');
            HASHLOOKUP.DEFINEDATA('MEANING');
            HASHLOOKUP.DEFINEDONE();
            CALL MISSING(ACRONYM, MEANING);
        END;
        SET DETAIL;
        RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
        IF RC = 0 THEN TOMEANING = MEANING;
        RUN;

        PROC PRINT DATA = RESULTS;
        RUN;
```

Out[5]:

The SAS System

Obs	DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
1	2006-12-21	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
2	2006-12-21	407	324-6674	BEG	TU	101	BIG EVIL GRIN
3	2006-12-21	407	312-9088	BFN	SYS	101	BYE FOR NOW
4	2006-12-21	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
5	2006-12-22	407	312-9088	C&G	AFAIK	103	

/ HASH ERROR1/

```
IF 0 THEN SET LOOKUP_1;
```

In [6]:
```
DATA _R1;
/* IF 0 THEN SET LOOKUP_1; */
DROP ACRONYM MEANING RC;
IF _N_ = 1 THEN DO;
    DECLARE HASH HASHLOOKUP(DATASET:'LOOKUP_1');
    HASHLOOKUP.DEFINEKEY('ACRONYM');
    HASHLOOKUP.DEFINEDATA('MEANING');
    HASHLOOKUP.DEFINEDONE();
    CALL MISSING(ACRONYM, MEANING);
END;
SET DETAIL;
RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
IF RC = 0 THEN TOMEANING = MEANING;
RUN;

PROC PRINT DATA = _R1;
RUN;
```

Out[6]:
```
174  ods listing close;ods html5 (id=saspy_internal) file=stdout options(bitmap_mode='inline') devi
ce=svg style=HTMLBlue; ods
174! graphics on / outputfmt=png;
NOTE: Writing HTML5(SASPY_INTERNAL) Body file: STDOUT
175
176  DATA _R1;
177  /* IF 0 THEN SET LOOKUP_1; */
178  DROP ACRONYM MEANING RC;
179  IF _N_ = 1 THEN DO;
180      DECLARE HASH HASHLOOKUP(DATASET:'LOOKUP_1');
181      HASHLOOKUP.DEFINEKEY('ACRONYM');
182      HASHLOOKUP.DEFINEDATA('MEANING');
183      HASHLOOKUP.DEFINEDONE();
184      CALL MISSING(ACRONYM, MEANING);
185  END;
186  SET DETAIL;
187  RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
188  IF RC = 0 THEN TOMEANING = MEANING;
189  RUN;
ERROR: Type mismatch for key variable ACRONYM at line 183 column 5.
ERROR: Hash data set load failed at line 183 column 5.
ERROR: DATA STEP Component Object failure.  Aborted during the EXECUTION phase.
NOTE: The SAS System stopped processing this step because of errors.
WARNING: The data set WORK._R1 may be incomplete.  When this step was stopped there were 0 observat
ions and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.02 seconds

190
191  PROC PRINT DATA = _R1;
192  RUN;
NOTE: No observations in data set WORK._R1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

193
194  ods html5 (id=saspy_internal) close;ods listing;

195
```

/ HASH ERROR2/

```
CALL MISSING(ACRONYM, MEANING);
```

In [7]:
```
/*
    HASH ERROR2
*/
DATA _R2;
IF 0 THEN SET LOOKUP_1;
DROP ACRONYM MEANING RC;
IF _N_ = 1 THEN DO;
    DECLARE HASH HASHLOOKUP(DATASET:'LOOKUP_1');
    HASHLOOKUP.DEFINEKEY('ACRONYM');
    HASHLOOKUP.DEFINEDATA('MEANING');
    HASHLOOKUP.DEFINEDONE();
/*    CALL MISSING(ACRONYM, MEANING); */
END;
SET DETAIL;
RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
IF RC = 0 THEN TOMEANING = MEANING;
RUN;

PROC PRINT DATA = _R2;
RUN;
```

Out[7]:

The SAS System

Obs	DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
1	2006-12-21	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
2	2006-12-21	407	324-6674	BEG	TU	101	BIG EVIL GRIN
3	2006-12-21	407	312-9088	BFN	SYS	101	BYE FOR NOW
4	2006-12-21	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
5	2006-12-22	407	312-9088	C&G	AFAIK	103	

/ *HASH ERROR3*/

```
RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
IF RC = 0 THEN TOMEANING = MEANING;
```

```
In [8]:  /*
             HASH ERROR3
         */
         DATA _R3;
         IF 0 THEN SET LOOKUP_1;
         DROP ACRONYM MEANING RC;
         IF _N_ = 1 THEN DO;
             DECLARE HASH HASHLOOKUP(DATASET:'LOOKUP_1');
             HASHLOOKUP.DEFINEKEY('ACRONYM');
             HASHLOOKUP.DEFINEDATA('MEANING');
             HASHLOOKUP.DEFINEDONE();
             CALL MISSING(ACRONYM, MEANING);
         END;
         SET DETAIL;
         RC = HASHLOOKUP.FIND(KEY:TOACRONYM);
         TOMEANING = MEANING;
         RUN;

         PROC PRINT DATA = _R3;
         RUN;
```

Out[8]:

The SAS System

Obs	DATE	AREACODE	PHONENUM	TOACRONYM	FROMACRONYM	SECRETCODE	TOMEANING
1	2006-12-21	407	312-9088	AFAIK	DQMOT	103	AS FAR AS I KNOW
2	2006-12-21	407	324-6674	BEG	TU	101	BIG EVIL GRIN
3	2006-12-21	407	312-9088	BFN	SYS	101	BYE FOR NOW
4	2006-12-21	407	312-9088	BTDT	IHU	102	BEEN THERE, DONE THAT
5	2006-12-22	407	312-9088	C&G	AFAIK	103	BEEN THERE, DONE THAT

In [ ]: