```
$SOIF{„Archive"}  =~ s|/TR/Data/|$ftpURL/| if $SOIF{„Archive"};
$SOIF{„Document"} =~ s|/TR/Data/|$ftpURL/| if $SOIF{„Document"};

while (($k, $v) = each %SOIF) {
        &soif'print_item($k, $v);
}
exit 0;
```

A.2 broker/query.html

A.3 broker/BrokerQuery.pl.cgi

A.4 broker/lib/BrokerQuery.cf

A.5 broker/idx2html.cgi

```perl
            $SOIF{„$what“}=$1; }
        if(/^Document-ID:\s+(.*)$/){                              $what=„Document-ID“;
         $SOIF{„$what“}=$1; }
    }


    #is there an archive file?
    if($file=~m/^(.*).$IDX$/){
        local($pref)=$1;

        $debug && print „pref=$pref\n“;

        $f=“${pref}.lha“;        if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.zip“;        if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.tar“;        if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.tar.Z“;      if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.tar.gz“;     if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.tar.z“;      if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.gz“;         if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.Z“;          if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.z“;          if(-f $f){ $SOIF{„Archive“}=$f; }
        $f=“${pref}.lzh”;        if(-f $f){ $SOIF{„Archive“}=$f; }

        $debug && print „$pref.$IDX -> „,$SOIF{„Archive“},“\n“;
        $debug && print „\n“;
        $debug && print „\n“;
    }


    # Abstract
    while(<IN>){
        $SOIF{„Abstract“}.=$_;
    }
    close(IN);


    # Div. Infos aus Filesystem:
    #  Semester, Version, Document, Archive, Document-ID
    if(($fb,$type,$sem,$num)=$file=~m|.*/TR/(\S+)/(\S+)/(\S+)/(\S+).$IDX|){
        $debug && print „fb = $fb\n“;
        $debug && print „type = $type\n“;
        $debug && print „sem = $sem\n“;
        $debug && print „num = $num\n“;
        $debug && print „\n\n“;

        $did=$file;
        $did=~s,^/TR/Data/TR/,,;
        $did=~s,/,-,g;
        $did=~s,.idx,,;

        $debug && print „did=$did\n\n“;
        $SOIF{„Document-ID“}=$did;

        $SOIF{„Semester“}=$sem if !$SOIF{„Semester“};
        if(!$SOIF{„Vers“}){
          if($type eq „SC“){
              $SOIF{„Vers“}=$sem;
          }
        }
        $SOIF{„Typ“}=$type      if !$SOIF{„Typ“};
    }
```

A Files

A.1 gatherer/bin/IDX.sum

```perl
#!/soft/local/bin/perl5

# Felder:
#  Titel:                  (mandatory)
#  Keywords:               (mandatory)
#  Autor:                  (mandatory)
#  Betreuer:               (mandatory)
#  Semester:               (recommended)
#  Vers:                   (recommended)
#  Document:               (optional)
#  Archive:                (optional)
#  Document-ID:            (optional)
#  Typ:                    (recommended)
#
#  Abstract:               (mandatory)

$IDX="idx";# idx?
$debug = 0;
$ftpURL = „ftp://rfhs0003.fh.uni-regensburg.de";


unshift(@INC, „$ENV{‚HARVEST_HOME'}/lib");        # use local files
require „soif.pl";

#open(L,">>/tmp/IDX.sum.log") ;
#print L „$0: „,@ARGV,"\n";
#close(L);

$file = shift(@ARGV);
open(IN, „$file") || die „IDX.sum: Cannot read $file.\n";

while(<IN>){
        last if (/^\s*\n$/o);
      if( /^\s+(.*)$/){
          # Continuation line
          $SOIF{„$what"}.=" $1";
      }
      if(/^Titel:\s+(.*)$/){        $what="Titel";        $SOIF{„$what"}=$1;
       }
      if(/^Keywords:\s+(.*)$/){                          $what="Keywords";
       $SOIF{„$what"}=$1; }
      if(/^Autor:\s+(.*)$/){        $what="Autor";        $SOIF{„$what"}=$1;
       }
      if(/^Betreuer:\s+(.*)$/){                          $what="Betreuer";
       $SOIF{„$what"}=$1; }
      if(/^Typ:\s+(.*)$/){          $what="Typ";          $SOIF{„$what"}=$1;
       }
      if(/^Semester:\s+(.*)$/){                          $what="Semester";
       $SOIF{„$what"}=$1; }
      if(/^Version:\s+(.*)$/){      $what="Vers";         $SOIF{„$what"}=$1;
       }
      if(/^Document:\s+(.*)$/){                          $what="Document";
       $SOIF{„$what"}=$1; }
      if(/^Archive:\s+(.*)$/){                           $what="Archive";
```

4.2 BrokerQuery.pl.cgi

Tasks:

- accept fields from Form (query.html)
- form query to broker
- call the right BrokerQuery.cf for output-formatting

4.3 lib/BrokerQuery.cf

Changes:

- minor cosmetics
- make IDX-files accessible through calls to `idx2html.cgi`.

4.4 idx2html.cgi

Display given IDX-file and hilight the Dokument- and Archive-fields (if available) as hyper-links.

See appendix for a full listing.

> *<CR>*
- Is this information correct? [yes]:
  > *<CR>*
- Should I build this Broker now? [yes]:
  > *<CR>*
- Would you like to add a collection point to the Broker now? [yes]:
  > *<CR>*
- Enter the host from which to collect [rfhs0003]:
  > *rfhs0003.fh.uni-regensburg.de*
- Enter the port from which to collect [8500]:
  > *8900*
- Enter the collection type [3]:
  > *1*
- Is this information correct? [yes]:
  > *<CR>*
- Would you like to add a collection point to the Broker now? [no]:
  > *<CR>*

If the TR-related files should be stored under /TR/Harvest, then /soft/Harvest-1.4pl1/brokers/TR has to be copied to /TR/Harvest/broker:

```
% rm -fr /TR/Harvest/broker
% cd /soft/Harvest-1.4pl1/brokers/TR
% mkdir -p /TR/Harvest/broker
% tar cf - . | ( cd /TR/Harvest/broker ; tar vxf - )
% cd ..
% rm -fr TR
% ln -s /TR/Harvest/broker TR
```

- Directory: /TR/Harvest/broker/TR
- BrokerQuery.pl.cgi: copy from /soft/Harvest-1.4pl1/cgi-bin/BrokerQuery.pl.cgi and modify as described below.
- lib/BrokerQuery.cf: copy from /soft/Harvest-1.4pl1/cgi-bin/lib/BrokerQuery.cf and modify as rescribed below.

4.1 Query-Form (query.html)

The following fields can be searched after checking them:

- Titel (check)   -> search_titel
- Keywords (check)   -> search_kwds
- Autor (check)   -> search_autor
- Betreuer (check)   -> search_betr
- Document-ID (check)   -> search_docid
- Semester (check)   -> search_sem
- Typ: multiple choice: DA, SC, REF, PR, any   -> search_typ
- Fachbereich: IN, ... any   -> search_fb

See below for a full listing of the form.

3.3 Summarizers

- Directory: /TR/Harvest/gatherer/bin
- Abstract.sum:
  ```
  #!/bin/sh
  head -50 „$1“ | wrapit „Abstract“
  ```
- Empty.sum:
  ```
  #!/bin/sh
  ```
- SymbolicLink.sum:
  ```
  #!/bin/sh
  ```
- IDX.sum:
  ```
  <too big, see appendix:->
  ```

4. Broker

- If there's only a ‚summary.html' in /TR/Harvest/broker/TR, the broker must be re-created by running `/soft/Harvest-1.4pl1/bin/CreateBroker` with the following parameters:

  - Enter a one-line descriptive name for your Broker [No_Default]:
    > Technical Report Broker
  - Enter a one-word name for your Broker [No_Default]:
    > TR
  - Enter the hostname on which the broker will run [rfhs0003]:
    > rfhs0003.fh.uni-regensburg.de
  - Enter the port number on which the broker will run [8501]:
    > 8901
  - Enter the password to the broker administrative interface []:
    > geheim
  - Enter the hostname for your HTTP server. If you want to use a different port number than 80, append it to the hostname with a colon (e.g., www.cs.colorado.edu:9000) [rfhs0003]:
    > rfhs0003.fh.uni-regensburg.de
  - Enter your email address [feyrer@rfhs0003]:
    > hubert.feyrer@rz.uni-regensburg.de
  - Enter the name of the attribute that will be displayed to the user as one-line object description in search results [description]:
    > <CR>
  - Do you want to use Glimpse or WAIS as an Indexer [glimpse]:
    > glimpse
  - Enter the pathname to the glimpse command [/soft/Harvest-1.4pl1/lib/broker/glimpse]:
    > <CR>
  - Enter the pathname to the glimpseindex command [/soft/Harvest-1.4pl1/lib/broker/glimpseindex]:
    > <CR>
  - Enter the pathname to the glimpseserver command [/soft/Harvest-1.4pl1/lib/broker/glimpseserver]:
    > <CR>
  - Enter the host on which to run glimpseserver [rfhs0003]:
    > rfhs0003.fh.uni-regensburg.de
  - Enter the destination Broker directory [/soft/Harvest-1.4pl1/brokers/TR]:

ACHTUNG!!! Der Server läuft noch nicht, darum ist dieses Dokument auch noch nicht 100% fertig. Unter http://rfhs0003.fh.uni-regensburg.de/Harvest/brokers/TR/summary.html kann das bisherige Ergebnis betrachtet werden. aber Vorsicht: Baustelle! Der TR-Broker basiert auf einem Broker, der den NetBSD/amiga-Teil des Regensburger FTP-Servers indiziert und unter http://rfhs0003.fh.uni-regensburg.de/Harvest/brokers/netbsd/query.html zugängig ist.

Setup of the Harvest Brokers for the Technical Report Server of the FH Regensburg

Hubert Feyrer, 6.3.1996

1. Preparation

- `mkdir -p /TR/Harvest/gatherer`
- `mkdir -p /TR/Harvest/broker`
- `ln -s /TR/Harvest/gatherer /soft/Harvest-1.4pl1/gatherers/TR`
- `ln -s /TR/Harvest/brokers /soft/Harvest-1.4pl1/brokers/TR`

2. RunHarvest

- Path of gatherer: *t/soft/Harvest-1.4pl2/gatherers/TR*
- Port of gatherer: *8900*
- Path of broker: *t/soft/Harvest-1.4pl2/brokers/TR*
- Port of broker: *8901*

3. Gatherer

3.1 Specify workload
- Directory: /TR/Harvest/gatherer
- TR.cf:
  ```
  <RootNodes>
  file://localhost/TR/Data/TR/ URL=99999999,TR-files
  </RootNodes>
  ```
- TR-files:
  ```
  Allow  \.idx
  Deny .*
  ```

3.2 Customize type recognition
- Directory: /TR/Harvest/gatherer/lib
- `cp /dev/null bycontent.cf`
- `cp /dev/null byname.cf`
- byurl.cf:
  ```
  IDX                    ^.*\.idx$
  ```
- `cp /dev/null quick-sum.cf`
- stoplist.cf:
  ```
  Unknown
  ```