A Format for E-mailing Bibliographic Records

Status of this Memo

   This memo provides information for the Internet community.
   It does not specify an Internet standard. Distribution of
   this memo is unlimited.

Abstract

   This memo defines a format for E-mailing bibliographic records of
   technical reports. It is intended to accelerate the dissemination
   of information about new Computer Science Technical Reports (CS-TR).

INTRODUCTION

   Many Computer Science R&D organizations routinely announce new
   technical reports by mailing (via the postal services) the
   bibliographic records of these reports.

   These mailings have non-trivial cost and delay. In addition, their
   recipients cannot conveniently file them, electronically, for later
   retrieval and searches.

   Therefore, it is suggested that the publishing organizations would
   e-mail these announcements by using the following format.

   Organizations may automate to any degree (or not at all) both the
   creation of these records (about their own publications) and the
   handling of the records received from other organizations.

   This format is designed to be simple, for people and for machines,
   to be easy to read ("human readable") and create without any special
   programs, and to be compatible with E-mail.

   This format defines how bibliographic records are to be transmitted.
   It does not define what to do with them when received.

   This format is a "tagged" format with self-explaining alphabetic
   tags. It should be possible to prepare and to read bibliographic
   records using any text editor, without any special programs.

This format was developed with considerable help and involvement of Computer Science and Library personnel from several organizations, including CMU, CNRI, Cornell, ISI, Meridian, MIT, Stanford, and UC. Key contributions were provided by Jerry Saltzer of MIT, and Larry Lannom of Meridian. The initial draft was prepared by Danny Cohen and Larry Miller of ISI.

The use of this format is encouraged. There are no limitations on its use.

THE INFORMATION FIELDS

The various fields should follow the format described below.

<M> means Mandatory; a record without it is invalid.
<O> means Optional.

The tags (aka Field-IDs) are shown in upper case.

```
<M>  BIB-VERSION of this bibliographic records format
<M>  ID
<M>  ENTRY date
<O>  ORGANIZATION
<O>  TITLE
<O>  TYPE
<O>  REVISION
<O>  AUTHOR
<O>  CORP-AUTHOR
<O>  CONTACT for the author(s)
<O>  DATE of publication
<O>  PAGES count
<O>  COPYRIGHT, permissions and disclaimers
<O>  RETRIEVAL information
<O>  CR-CATEGORY
<O>  PERIOD
<O>  SERIES
<O>  FUNDING organization(s)
<O>  MONITORING organization(s)
<O>  CONTRACT number(s)
<O>  GRANT number(s)
<O>  LANGUAGE name
<O>  NOTES
<O>  ABSTRACT
<M>  END
```

META FORMAT
----

   * Keep It Simple.

   * One bibliographic record for each publication, where a
     "publication" is whatever the publishing institution defines
     as such.

   * A record contains several fields.

   * Each field starts with its tag (aka the field-ID) which is a
     reserved identifier (containing no separators) at the beginning
     of a new line with or without spaces before it), followed by two
     colons ("::"), followed by the field data.

   * Continuation lines: Lines are limited to 79 characters. When
     needed, fields may continue over several lines, with an implied
     space in between. In order to simplify the use no special marking
     is used to indicate continuation line. Hence, fields are
     terminated by a line that starts (apart from white space) with
     a word followed by two colons. Except for the "END::" that is
     terminated by the end of line.) For improved human readability
     it is suggested to start continuation lines with some spaces.

   * Several fields are mandatory and must appear in the record. All
     fields (unless specifically not permitted to) may be in any order
     and may be repeated as needed (e.g., the AUTHOR field). The order
     of the repeated fields is always preserved.

   * Only printable ASCII characters may be used. Hence, the
     permissible characters are ASCII codes 040 (Space) through 176(~)
     and line breaks which are \012 (LF) or \012\015 (CRLF). Empty
     lines indicate paragraph break. \009 (tab) must be replaced by
     spaces before submission. This specifically forbids tabs, null
     characters, DEL, backspaces, etc. (i.e., if used, the record is
     invalid.)

   Throughout this document the word "publisher" means the publishing
   organization of a report (e.g., a university or a department
   thereof), not necessarily an organization authorized to issue ISBN
   numbers.

                              EXAMPLE

     BIB-VERSION:: CS-TR-v2.0
              ID:: OUKS//CS-TR-91-123
           ENTRY:: January 15, 1992
    ORGANIZATION:: Oceanview University, Kansas, Computer Science
           TITLE:: The Computerization of Oceanview with High
                   Speed Fiber Optics Communication
            TYPE:: Technical Report
        REVISION:: 2, FTP retrieval information added
          AUTHOR:: Finnegan, James A.
         CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ, Oceanview,
                   KS 54321 Tel: 913-456-7890 <Finnegan@cs.ouks.edu>
          AUTHOR:: Pooh, Winnie The
         CONTACT:: 100 Aker Wood
            DATE:: December 1991
           PAGES:: 48
       COPYRIGHT:: Copyright for the report (c) 1991, by J. A. Finnegan.
                   All rights reserved. Permission is granted for any
                   academic use of the report.
       RETRIEVAL:: For full text with color pictures send a self-addressed
                   stamped envelope to Prof. J. A. Finnegan, CS Dept,
                   Oceanview University, Oceanview, KS 54321.
       RETRIEVAL:: ASCII available via FTP from JUPITER.CS.OUKS.EDU with the
                   pathname PUBS/computerization.txt. Login with FTP,
                   username ANONYMOUS and password GUEST.
                   File size: 123,456 characters
     CR-CATEGORY:: D.0
     CR-CATEGORY:: C.2.2 Computer Sys Org, Communication nets, Net Protocols
          SERIES:: Communication
         FUNDING:: FAS
        CONTRACT:: FAS-91-C-1234
      MONITORING:: FNBO
        LANGUAGE:: English
           NOTES:: This report is the full version of the paper with the
                   same title in IEEE Trans ASSP Dec 1976

    ABSTRACT::

    Many alchemists in the country work on important fusion problems.
    All of them cooperate and interact with each other through the
    scientific literature. This scientific communication methodology
    has many advantages. Timeliness is not one of them.

    END:: OUKS//CS-TR-91-123
    -------------------------- End of Example --------------------------

       For reference, the above example has about 1,750 characters (220
       words) including about 250 characters (40 words) in the abstract.

THE ACTUAL FORMAT

     In the following double-quotes indicate complete strings. They are
     included only for grouping and are not expected to be used in the
     actual records.

     The term "Open Ended Format" in the following means arbitrary text.

     The BIB-VERSION, ID, ENTRY, and END field must appear as the first,
     second, third, and last fields, and may not be repeated in the
     record. All other fields may be repeated as needed.

BIB-VERSION (M) -- This is the first field of any record. It is a
        mandatory field. It identifies the version of the format used
        to create this bibliographic record.

        BIB-VERSIONs that start with the letter X (case independent)
        are considered experimental. Bib-records sent with such a
        BIB-VERSION should NOT be incorporated in the permanent database
        of the recipient.

        Using this version of this format, this field is always:

        Format: BIB-VERSION:: CS-TR-v2.0

ID (M) -- This is the second field of any record. It is also a
        mandatory field. Its format is "ID:: XXX//YYY", where XXX is
        the publisher-ID (the controlled symbol of the publisher)
        and YYY is the ID (e.g., report number) of the publication as
        assigned by the publisher. This ID is typically printed on
        the cover, and may contain slashes.

        The organization symbols "DUMMY" and "TEST" (case independent)
        and any organization symbol starting with <X> (case
        independent) are reserved for test records that should NOT
        be incorporated in the permanent database of the recipients.

        Format: ID:: <publisher-ID>//<free-text>

        Example: ID:: OUKS//CS-TR-91-123

            **** See the note at the end regarding the ****
            **** controlled symbols of the publishers *****

ENTRY (M) -- This is a mandatory field. It is the date of creating this
        bibliographic record.

        The format for ENTRY date is "Month Day, Year". The month must
        be alphabetic (spelled out).   The "Day" is a 1- or 2-digit
        number. The "Year" is a 4-digit number.

        Format: ENTRY:: <date>

        Example: ENTRY:: January 15, 1992

ORGANIZATION (O) -- It is the full name spelled out (no acronyms,
        please) of the publishing organization. The use of this name
        is controlled together with the controlled symbol of the
        publisher (as discussed above for the ID field).

        Avoid acronyms because there are many common acronyms, such as
        ISI and USC. Please provide it in ascending order, such as
        "X University, Y Department" (not "Y Department, X University").

        Format: ORGANIZATION:: <free-text>

        Example: ORGANIZATION:: Stanford University, Computer Science

TITLE (O) -- This is the title of the work as assigned by the author.
        This field should include the complete title with all the
        subtitles, if any.

        If the publication has no title (e.g., in withdrawal), a blank
        TITLE field should be included.

        Format: TITLE:: <free-text>

        Example: TITLE:: The Computerization of Oceanview with High
                              Speed Fiber Optics Communication

TYPE (O) -- Indicates the type of publication (summary, final project
        report, etc.) as assigned by the issuing organization.

        Format: TYPE:: <free-text>

        Example: TYPE:: Technical Report

REVISION (O) -- Indicates that the current bibliographic record is
        a revision of a previously issued record and is intended to
        replace it. Revision information consists of an integer
        followed by a comma, and by text in an open ended format.
        The revised bibliographic record should contain a complete
        record for the publication, not just a list of changes to
        the old record. The default assumption is that a record is
        not a revision (i.e., specify only if it is), with that integer
        being zero.

        The first token in this field is an integer revision number.
        Higher numbers indicate later revisions. Use the text to
        describe the revision. Reasons to send out a revised record
        include an error in the original, change in the retrieval
        information, or withdrawal (see below).

        Format: REVISION:: N, <free-text>

        Example: REVISION:: 2, FTP retrieval information added

    WITHDRAWING: A withdrawal of a record is a special case of revising
        it. Hence, the standard way to withdraw records is by sending a
        revision record with (at least) all the mandatory fields, and an
        optional explanation in the NOTES field.

        It is OK on withdrawal to eliminate the title, by not providing
        the TITLE field it or by providing it with no text (blank).

        Example for withdrawing a bibliographic record::

            BIB-VERSION:: CS-TR-v2.0
            ID::   OUKS//CS-TR-91-123
            ENTRY::   January 25, 1992
            ORGANIZATION:: Oceanview University, Kansas, Computer Science
            TITLE::
            REVISION::   4, withdrawn
            NOTES::   Withdrawn, found to be irrelevant
            END::   OUKS//CS-TR-91-123

        This new record will replace all the fields of the previous
        record for that publication. In this example it will eliminate
        the title, the retrieval information provided earlier, and not
        mention the authors.

AUTHOR (O) -- Personal names only. Normal last name first inversion.
        Editors should be listed here as well, identified with the
        usual "(ed.)" as shown below in the last example.

        If the report was not authored by a person (e.g., it was
        authored by a committee or a panel) use CORP-AUTHOR (see below)
        instead of AUTHOR.

        Multiple authors are entered by using multiple lines, each in
        the form of "AUTHOR:: <free-text>".

        The system preserves the order of the authors.

        Format: AUTHOR:: <free-text>

        Example: AUTHOR:: Finnegan, James A.
                  AUTHOR:: Pooh, Winnie The
                  AUTHOR:: Lastname, Firstname (ed.)

CORP-AUTHOR (O) -- The corporate author (e.g., a committee or a
        panel) that authored the report, which may be different from
        the ORGANIZATION issuing the report.

        In entering the corporate name please omit initial "the" or "a".
        If it is really part of the name, please invert it.

        Format: CORP-AUTHOR:: <free-text>

        Example: CORP-AUTHOR:: Committee on long-range computing

CONTACT (O) -- The contact for the author(s).
        Open-ended, most likely E-mail and postal addresses.

        You may provide a CONTACT field for each author separately,
        or for all the AUTHOR fields.

        E-mail addresses should always be in "pointy brackets"
        (as in the example below).

        Format: CONTACT:: <free-text>

        Example: CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview
                           Univ., Oceanview, Kansas, 54321
                           Tel: 913-456-7890 <Finnegan@cs.ouks.edu>

DATE (O) -- The publication date. The formats are "Month Year" and
          "Month Day, Year". The month must be alphabetic (spelled out).
          The "Day" is a 1- or 2-digit number. The "Year" is a 4-digit
          number.

          Format: DATE:: <date>

          Example: DATE:: January 1992
          Example: DATE:: January 15, 1992

PAGES (O) -- Total number of pages, without being too picky about it.
          Final numbered page is actually preferred, if it is a reasonable
          approximation to the total number of pages.

          Format: PAGES:: <number>

          Example: PAGES:: 48

COPYRIGHT (O) -- Copyright, permissions and disclaimers. Open
          ended format. The COPYRIGHT field applies to the cited
          report, rather than to the current bibliographic record.
          On advice of counsel it is suggested that you seek the
          advice of yours.

          Format: COPYRIGHT:: <free-text>

          Example: COPYRIGHT:: Copyright for the report (c) 1991,
                             by J. A. Finnegan. All rights reserved.
                             Permission is granted for any academic
                             use of the report.

RETRIEVAL INFORMATION (O) -- Open-ended format describing how to get
          a copy of the full text. It may include anything from FTP
          instructions to a variety of files (e.g., ASCII, TeX, and
          PostScript) to "Send $4.50 to ..." or "Send E-mail to <X@Y>".

          It is suggested to repeat this field for each retrieval option
          (e.g., one line for the FTP instructions to the ASCII version,
          and another for the PostScript version). When offering files
          like TeX all the related files (e.g., "\input mystyle") should
          be included. Please provide file sizes (in characters).

          Means are not defined yet for providing the information needed
          for automatic retrieval of files (such as via FTP). They are
          expected to be defined in the near future.

     No limitations are placed on the dissemination of the
     bibliographic records. If there are limitations on the
     dissemination of the publication, it should be protected by
     some means such as passwords. This format does not address
     this protection.

     Format: RETRIEVAL:: <free-text>

     Example: RETRIEVAL:: For full text with color pictures send
                         a self-addressed stamped envelope to
                         Prof. J. A. Finnegan, CS Dept,
                         Oceanview University, Oceanview, KS 54321.
              RETRIEVAL:: ASCII available via FTP from
                         JUPITER.CS.OUKS.EDU with the pathname
                         PUBS/computerization.txt.
                         Login with FTP, username ANONYMOUS and
                         password GUEST.
                         File size: 123,456 characters

CR-CATEGORY (O) -- Specify the CR-category. The CR-category (the
        Computer Reviews Category) index (e.g., "B.3") should always be
        included, optionally followed by the name of that category. If
        the name is specified it should be fully specified with parent
        levels as needed to clarify it, as in the second example below.
        Use multiple lines for multiple categories.

        The January 1992 issue of CR has the full list of these
        categories, with a detailed discussion of the CR Classification
        System, and a full index. Typically the full index appears in
        every January issue, and the top two levels in every issue.

        Format: CR-CATEGORY:: <free-text>

        Example: CR-CATEGORY:: D.1

        Example: CR-CATEGORY:: B.3 Hardware, Memory Structures

PERIOD (O) -- Time period covered (date range). Applicable primarily to
        progress reports, etc. Any format is acceptable, as long as the
        two dates are separated with " to " (the word "to" surrounded by
        spaces) and each date is in the format allowed for dates, as
        described above for the date field.

        Format: PERIOD:: <date> to <date>

        Example: PERIOD:: January 1990 to March 1990

SERIES (O) -- Series title, including volume number within series.
        Open-ended format, with producing institution strongly
        encouraged to be internally consistent.

        Format: SERIES:: <free-text>

        Example: SERIES:: Communication

FUNDING (O) -- The name(s) of the funding organization(s).

        Format: FUNDING:: <free-text>

        Example: FUNDING:: DARPA

MONITORING (O) -- The name(s) of the monitoring organization(s).

        Format: MONITORING:: <free-text>

        Example: MONITORING:: ONR

CONTRACT (O) -- The contract number(s).

        Format: CONTRACT:: <free-text>

        Example: CONTRACT:: MMA-90-23-456

GRANT (O) -- The grant number(s).

        Format: CONTRACT:: <free-text>

        Example: GRANT:: NASA-91-2345

LANGUAGE (O) -- The language in which the report is written.
        Please use the full English name of that language.

        Please include the Abstract in English, if possible.

        If the language is not specified, English is assumed.

        Format: LANGUAGE:: <free-text>

        Example: LANGUAGE:: English

NOTES (O) -- Miscellaneous free text.

        Format: NOTES:: <free-text>

        Example: NOTES:: This report is the full version of the paper
                          with the same title in IEEE Trans ASSP Dec
                          1976

ABSTRACT (O) -- Highly recommended, but not mandatory. Even though no
        limit is defined for its length, it is suggested not to expect
        applications to be able to handle more than 10,000 characters.

        The ABSTRACT is expected to be used for subject searching since
        titles are not enough. Even if the report is not in English, an
        English ABSTRACT is preferable. If no formal abstract appears
        on document, the producers of the bibliographic records are
        encouraged to use pieces of the introduction, first paragraph,
        etc.

        Format: ABSTRACT:: xxxx xxxxxxxx
                            xxxx xxxxxxxx

                            xxxx xxxxxxxx
                            xxxx xxxxxxxx

END (M) -- This is a mandatory field. It must be the last entry of a
        record, identifying the record that it ends, by stating the same
        ID that was used at the beginning of the records, in its "ID::".

        Format: END:: XXX

        Example: END:: OUKS//CS-TR-91-123

            >>>>>>>   [END OF FORMAT DEFINITION]   <<<<<<<

A Note Regarding the Controlled Symbols of the Publishers

   In order to avoid conflicts among the symbols of the publishing
   organizations (the XXX part of the "ID:: XXX//YYY") it is suggested
   that the various organizations that publish reports (such as
   universities, departments, and laboratories) register their
   <publisher-ID> symbols and names, in a way similar to the
   registration of other key parameters and names in the Internet.

   Danny Cohen <Cohen@ISI.EDU> of ISI, has agreed to coordinate this
   registration for the publishers of Computer Science technical
   reports. It is suggested that before using this format the
   publishing organizations would coordinate with him (by e-mail) their
   symbols and the names of their organizations. [Discussions are in
   progress to have these publisher-IDs registered with the Internet
   Assigned Numbers Authority (IANA) and listed in future editions of
   the Assigned Numbers document.]

   In order to help automated handling of the received bibliographic
   records, it is expected that the producers of bibliographic records
   will always use the same name, exactly, in the ORGANIZATION field.

Security Considerations

   Security issues are not discussed in this memo.

Author's Address

   Danny Cohen
   USC - Information Sciences Institute
   4676 Admiralty Way
   Marina del Rey, California 90292-6695

   Phone: 310-822-1511

   Fax: 310-823-6714

   EMail: Cohen@ISI.EDU