

記録類の全文入力の試み

—『石山本願寺日記』を素材に—

水 藤 真

はじめに

1. 全文入力に何を期待するか？
2. 入力作業の実例

3. 『石山本願寺日記』データベースの内容

4. コンピューターの効用のほど
5. 今後の課題

論文要旨

この「記録類の全文入力の試み」というのは、日記や編さん物の全文をコンピューターに入力して、その検索・統計などの用途に益ならしめようとするものである。例えば、「祭」という言葉が何という日記の何年何月何日条に出て来るかを、コンピューターを使って、瞬時に知ることは出来ないか、などという期待を実現出来ないかという試みである。

当然、過去に記された日記には色々な特徴があり、コンピューター化に適したものとそうでないものがある。例えば、日記に記された文章は、それぞれの日にちで長さが異なる。また、文章の途中に書き直したものや、後で追記されたものが挿入されるとか、また現在では使われていない文字が使われているとか、様々なことがあり、単純にただ文章をコンピューターに打ち込めば良いというものではない。この為の諸条件の整備が最初の難関であった。

次には、コンピューターの操作が難しいという問題があった。

こうして、いくつかの閑門を突破しつつ、次の努力目標を掲げた。

誰でもが使えるような易しいシステムであること。あらかじめキーワードを準備するのではなく、任意の文字列で検索出来ること。この2点である。こうしてシステムは兎も角動くようになった。しかし、一旦稼働することが分かると、次には新たな問題が持ちあがった。例えば検索の為の時間である。当初、時間はいくらかかっても、任意の文字列で検索出来ることを至上目標とした。当然の結果とはいえ、時間は40万字の検索で12、3分を要することとなった。それでも、初めて検索結果の出来た時は感動であった。しかし、2度、3度と異なる言葉を検索しようとした時、これではかなり障害となる。そこで今度は時間短縮を計ることが重要課題となつた。そして、今、1件当たり約1、2分で検索出来る。本稿はこの悪戦苦闘の記録である。