

技術情報の統合管理

2M-6

一文書のパラグラフ管理

永野靖忠, 七野 剛, 金近秀明, 平田孝雄, 前田 暁

三菱電機(株)情報電子研究所

1. はじめに

設計部門において取り扱われる各種仕様書や規格文書などの技術文書は非常に多く、その中に記された技術情報は様々な設計業務に参照、流用される重要な情報である。これらの技術情報をデータベースとして蓄積管理し、有効に利用できるシステムを実現することにより設計開発の効率化を図ろうとする要求がある。参照又は流用される技術情報は文書の一部に存在することが多いため、この様なシステムにおいては、文書単位の検索のみならず、文書の必要な部分の検索が要求される。

このような要求に対し、文書を書式情報にもとづいてパラグラフに分割し、パラグラフ単位の検索を可能にしたシステムの開発を行っている。そのシステムの構成、機能および実現方式について報告する。

2. システム構成

システムの構成は図1に示すごとく、ホスト計算機とそれに接続された複数のエンジニアリング・ワークステーション(EWS)で構成される。EWSにおいて画像、図形、テキスト情報を含む技術文書を作成する。作成された文書はネットワークに接続されたホスト計算機に送られる。ホスト計算機の文書ファイルシステムとリレーショナルデータベース(RDB)にそれぞれ文書データおよびその台帳データが格納され、パラグラフ情報もホスト計算機に蓄積される。各EWSよりネットワークによってホスト計算機に蓄積された文書を台帳やパラグラフ情報を用いて検索するシステム構成である。

3. 機能

3.1. 文書単位の検索機能: オフィスの文書保管形態をモデル化した電子ファイルのキャビネット/バイнда/ドキュメントという保管階層に対する検索条件またはRDBの台帳に対する検索条件により文書単位の検索を行うことができる。

3.2. パラグラフ単位の検索: 書式によって定められた項目や内容でパラグラフに対する検索条件を指定する。3.1. に示した文書単位の検索に必要な検索条件とパラグラフに対する検索条件とを組み合わせると、条件を満たす文書のパラグラフが出力される。この機能により例えば文書の「概要」や「機能」

一覧の出力など、必要なパラグラフのみを検索して目的の情報を得ることができる。

4. 実現方式

4.1. パラグラフ情報の抽出: パラグラフ単位の検索を可能にするためには文書をパラグラフに分割することとそのパラグラフに検索キーを付加することが必要である。この作業は人手による対話入力方式では膨大なコストがかかるため自動化されなければならない。そこで、当システムでは技術文書が定型書式で書かれることに着目し、書式情報を用いてパラグラフ情報の抽出を行うようにしている。即ち、書式情報として書式のパラグラフ構成とそのパラグラフに付加すべき検索キーの情報を設定しておく。書式情報を用いて、入力された文書とその書式のパラグラフ構成と比較して対応するパラグラフに分割し、検索キーを付加して、書式情報にもとづくパラグラフ情報を自動的に抽出する。

4.2. パラグラフ検索: パラグラフ情報を用いて検索を行う。パラグラフに付加されている検索キーは書式によって相違する。例えば「概要」と「要約」などは異なる検索キーが付加されるが、書式によっては同一の意味で使用されている。この様な同一の意味や意味的に包含関係にある検索キーは自動的に同一の検索キーとして処理され、書式の相違を意識しないでパラグラフに対する検索条件の指定を可能にしている。

5. おわりに

文書検索においてパラグラフ検索が可能なシステムについて述べたが、さらにパラグラフの内容に対する検索条件をどの程度指定できるかなどについては今後の課題である。

文献

1) 板橋他「オフィスシステムにおける電子キャビネットと仮想オフィス」情報処理学会第34回全国大会(1987)

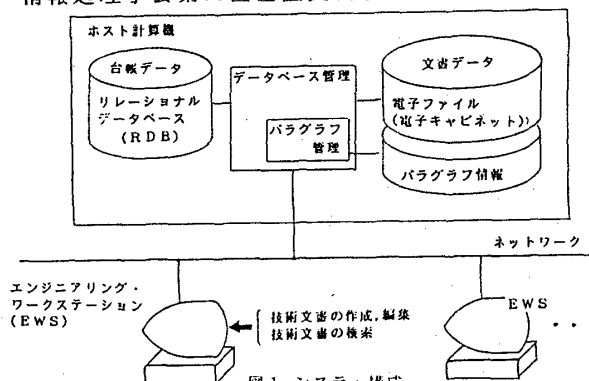


図1. システム構成

Integrated Management for Engineering Information
-Management for Engineering Document Paragraph-
Yasutada NAGANO, Go SHICHINO, Hideaki KANECHIKA,
Takao HIRATA, Akira MAEDA
Mitsubishi Electric Corp.