

検索構造と検索項目

吉武道子、吉原一紘

金属材料技術研究所

〒305 つくば市千現1-2-1

科学技術庁の省際ネットワーク構想とのリンクのもと表面分析研究会において作成する表面分析データベースについて、提供されている環境、現段階（5月8日）での状況、及び検索構造や検索項目について表面分析研究会において検討する問題を整理した。

1. はじめに

ネットワークを介したデータベースをインターネットの使えるUNIXベースで構築しようとする場合、Relational Data Base Management System (RDBMS)と呼ばれるデータベース管理環境上に作成するのが普通である。そしてユーザーとサーバをネットワークプログラムを介して結び、SQLと呼ばれるデータベースの標準言語をやり取りすることによ

って検索などを行う。以上の環境は既に提供されているものであって、我々はRDBMSとしてOracle7、ネットワークプログラムとしてVGUIDEを利用することにした。我々のデータベースではどのソフトを利用しても機能に大差はないのでシェアやアフターケアを参考に選んだ。ここでは、どこまでが提供された環境で、我々が検討しなければならないのは何か、について述べる。

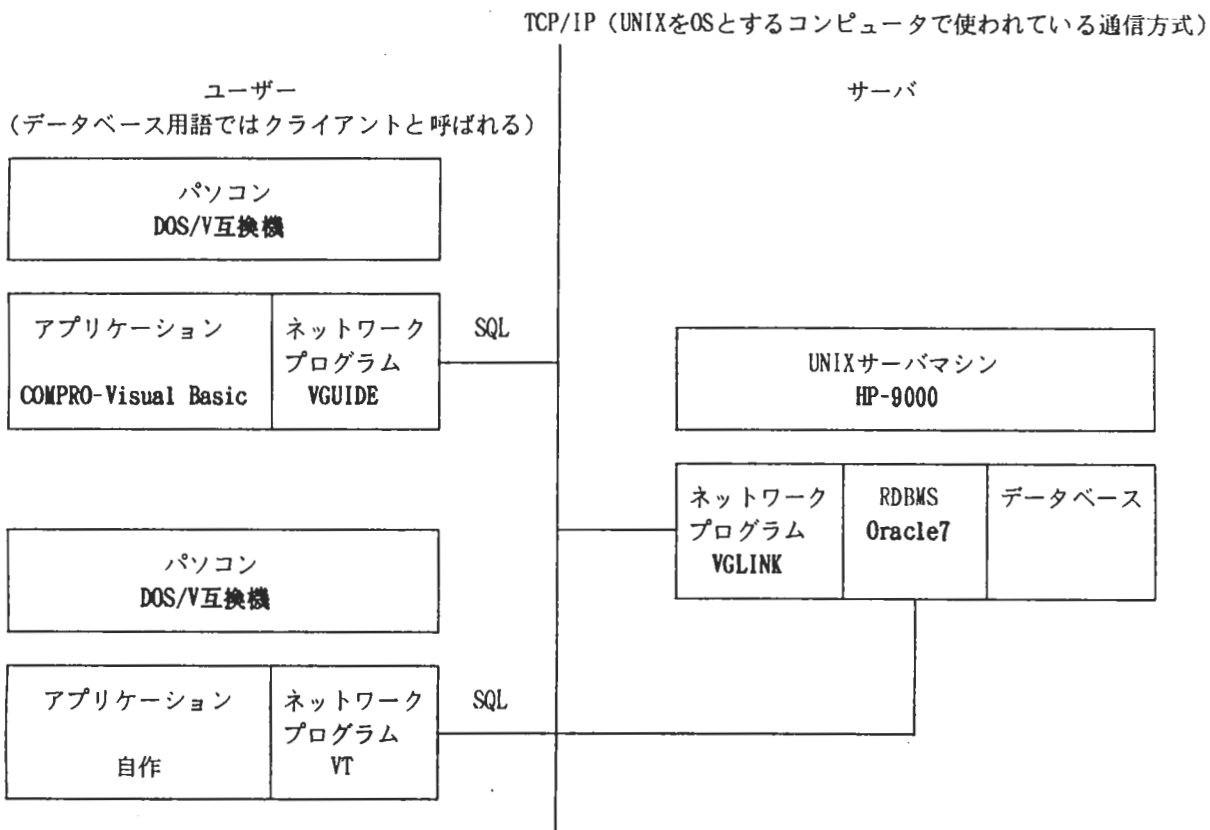


図1 ネットワークデータベースに提供されている環境の概要図

2. 提供されている環境の概要

ユーザーからサーバへアクセスする環境を簡単に示したのが図1である。図中、普通の文字で書かれているのは一般的呼称、太字は我々のシステムで採用したものを表す。ここでは、異なる装置からのデータを共通で扱うためのツールが含まれているCOMPROを用いてデータベースにアクセスする場合の検索を検討する。

VGUIDEやVGLINGはVisual Basic等のアプリケーションソフトからアクセスしやすいように作られたソフトウェアであり、UNIXマシンのコマンドとOracleの利用方法及びSQL言語の知識があれば、VGUIDEやVGLING、Visual Basicに縛られずに直接Oracle内に作られたデータベースにアクセスすることができる。また、UNIXマシンのコマンドの知識がなくても、Oracleの利用方法及びSQL言語及びVisual Basic

の知識（プログラミングができること）があれば、COMPROを使わなくてもデータベースにアクセスできる。図1に示した環境は、これらのCOMPROを使わないユーザーのアクセスを何ら妨げるものではない。

次に、Oracleのデータ管理方法（Oracleに限らず、UNIX用RDBMSではどれも同じ）について概念的な感じが得られるよう例を挙げる。ここに書いたようなOracle上での表（Master Indexと呼ぶことにする）はサーバ内に収まっていて、COMPROからアクセスするユーザーには見えない。ユーザーはあたかも図2に示されるような表がコンピュータ内にあり、そこから、例えば、Bの含まれたSiをAESで測定したSi LMMピークのファイル（スペクトル番号で一義的に決定される）にはどのようなものがあるか、いくつあるか、をリストアップするような作業を行うことになる。

<ヘッダーテーブル>

スペクトル番号	機関名	測定者	測定年	装置メーカー	測定手法	励起源の種類	試料名	..
0000121	金材研	吉武道子	1994	ファイ	XPS	Mg	Au	..
0000122	電総研	一村信吾	1995	日本電子	AES	e-gun	Si	..
0000123	物質研	小島勇夫	1995	VG	AES	e-gun	MgO	..
...

行

列
(カラム)

<元素テーブル>

スペクトル番号	元素の含まれ方	H	He	Li	Be	B	C	N	O	F	Ne	Na	..
0000121	detected						1.2		2.3				..
0000121	known impurity											0.0001	..
0000122	known impurity				0.003								..
...

図2 Oracle上での表の例

3. サーバ上のMaster Index

COMPROから検索する場合、検索したい項目はサーバ上のMaster Indexのカラム名に含まれていなければならないし、第4章の検索イメージのところ述べるような検索結果の表示の一覧で表示させたい項目もこの表のカラム名に含まれていなければ

ならない。どのような項目をMaster Indexのカラムに含めるかについて4月のデータベース委員会での議論を基に検討した結果、表1に示すMaster Indexを作成することにした。ただし、このMaster Indexは最終版ではなく、データベースの作り込み過程で若干の変更もあることに留意していただきたい。

表1 データベースMaster Indexカラム名

Name	Character	Example
FileId	Char 20	00000104
InstId	Char 20	NRIM
ModId	Char 20	PHI-660
OprtId	Char 20	Yoshihara
ExprId	Char 80	quantify
Commnt	Char 80	Round robin
ExpMod	Char 20	NORM
NumRgn	number(4)	2 (After registration, NPL file - 1)
NumBlk	number(8)	3 (After registration, NPL file - 1)
BlckId	Char 20	4th block id (After registration, NPL file 1st block)
SmpId	Char 20	1st sample id
YearIF	number(4)	1991
MonthY	number(2)	4
DayMnt	number(2)	12
DHPProc	Char 80	differentiated
ErgScl	Char 80	C ("C" means "this machine is calibrated by the Cu***)
IntScl	Char 80	A: Cu spectrum(Num:*****) is attached
ResScl	Char 80	none
HostMt	Char 20	stainless steel
IUPACn	Char 20	unkown
CASNum	Char 20	0 (unknown case)
HostCm	Char 20	Fe 74 Cr 18 Ni 8
BlkPrt	number(5, 2)	99.9
Strctr	Char 20	fcc
FormPr	Char 20	Roof
Crystn	Char 20	P
MatFam	Char 20	M
MatCls	Char 20	C
ExPrep	Char 80	polish
InPrep	Char 80	ion sputtering
SmpTmp	number(6, 2)	298
ChgCnt	Char 80	none
ElmExt	Char 80	Fe-Co-Ni
SmpInf	Char 80	supplied by NRIM
SpcInf	Char 80	AF0045RT 4th of 6 blocks (original name and blocks)
SpcGrp	Char 80	00000101-00000106(Group spectra)
SpcRnk	Char 20	R
Technq	Char 20	AES dir
AnlSrc	Char 20	electron gun
AnlErg	number(8, 2)	5000
TakOff	number(5, 2)	0
SpcLbl	Char 20	Fe
TrnLbi	Char 20	LMM
AbsLbl	Char 20	kinetic energy
AbsStr	number(8, 2)	0
AbsEnd	number(8, 2)	1200

表2 データベースMaster Indexカラム名の意味するもの

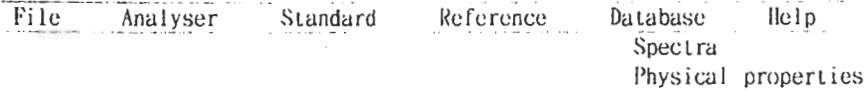
Name	Meaning of 'Name'	Example
FileId	File Identifier	00000104
InstId	Institution Identifier	NRIM
ModlId	Instrument Model Identifier	PHI-660
OprId	Operator Identifier	Yoshihara
ExprId	Experiment Identifier	quantify
Commnt	Comment	Round robin
ExpMod	Experiment Mode	NORM
NumRgn	Number of Spectral Regions	2 (After registration, NPL file → 1)
NumBlk	Number of Blocks	3 (" " , NPL file → 1)
BlckId	Block Identifier	4th block id (" " , NPL file → 1st block)
SmplId	Sample Identifier	1st sample id
YearIF	Year in full	1991
MonthY	Month	4
DayMnt	Day of Month	12
DHProc	Data Handling Procedure	differentiated
ErgScl	Energy Scale Calibration Procedure	C ("C" means "calbrated by the Cu***")
IntScl	Intensity Scale Calibration Procedure	A: Cu spectrum(Num:*****) is attached
ResScl	Energy Resolution Calibration Procere	none
HostMt	Host Material (通称)	stainless steel
IUPACn	IUPAC Chemical Name	unkown
CASNum	Chemical Abstracts Registry Number	0 (unknown case)
HostCm	Host Material Composition	Fe 74 Cr 18 Ni 8
BlkPrt	Bulk Purity	99.9
Stretr	Structure	fcc
FormPr	Form of Products	Roof
Crystn	Crystallinity	P
MatFam	Material Family	M
MatCls	Special Material Classes	C
ExPrep	Ex situ Preparation	polish
InPrep	In situ Preparation	ion sputtering
SmpTmp	Specimen Temperature	298
ChgCnt	Charge Control Conditions	none
ElmExt	Elements Existing on the Surface	Fe-Co-Ni
SmpInf	Specimen Information	supplied by NRIM
SpcInf	Spectrum Information	AF0045RT 4th of 6 blocks (original name and blocks)
SpcGrp	Spectrum Group	00000101-00000106(Group spectra)
SpcRnk	Spectrum Ranking	R
Technq	Technique	AES dir
AnlSrc	Analysis Source Label	electron gun
AnlErg	AnalysisSource Characteristic Energy	5000
TakOff	Analyser Axis Take Off Polar Angle	0
Spclbl	Species Label	Fe
TrnLbl	Transition Label	LMM
AbsLbl	Abssissa Label	kinetic energy
AbsStr	Abssissa Start	0
AbsEnd	Abssissa End	1200

COMPROから検索する場合はMaster Indexのカラム名を知っている必要はなく、COMPROの画面上でそれぞれのカラム名に対応するボタンをクリックすればよいようにしておく。

COMPROを使わずに検索する場合には、このMaster Indexのカラム名をタイプインして検索することになるので、表1のNameとその意味するところ(表2)を正確に把握している必要がある。

3. 検索構造と検索項目のイメージ

(1) まず、メニューバーの Database の下に、スペクトルか物理定数かどちらを検索するかというプルダウンメニューがでる。



(2) Spectra を選択すると、下のような検索項目入力画面がでる。

General	Measurement	Process&Calibration	Sample
Institute	Instrument manufacturer	Processed	Name
Operator	Technique	Energy scale calib	Detected element
Year	Source type	Intensity scale calib	Sample comment
Comment	Scan energy range		

select

input

AND OR NOT SEARCH CLEAR

金材研で1994年以降に取得したFeのFe 2p XPSスペクトルを検索したいとき：

Institute をクリック、 をクリック、input の中に NRIM と入力、 をクリック、Year をクリック、 をクリック、1994 を入力、 をクリック、Technique をクリック、 をクリック、XPS を入力、 をクリック、Name をクリック、 をクリック、Fe を入力、 をクリック、Scan energy range をクリック、 をクリック、710 を入力、 をクリック。

検索に該当するファイル名と、測定や試料に関する情報が一覧表となって示される。

ファイル番号	測定機関 サンプルコメント	測定年	手法 スペクトルグループ	励起源	試料名	エネルギー範囲	データ処理	検出元素
00104826	NRIM as-received	1994	XPS 00104826-00104832	Mg	Fe	740-700	intensity calib	Fe, O, C
00018270	NRIM heated at 800K	1995	XPS 00018250-00018274	Al	Fe	760-700		Fe, S
00021932	NRIM cleaned	1995	XPS 00021930-00021945	Mg	Fe	754-700	sno 7	Fe

4. 検討すべき課題

以下は、早急に検討し、省際ネットワークのプロジェクト中に試行錯誤を終了する必要がある課題である。

(1) 各自が検索にどのようなイメージを持っているかを明確にする。図書検索などすでに利用している情報検索システムなどを参考に、測定装置や測定機関を問わずスペクトルが利用できる場合、どのような検索をしたいか、それにはどのような手順になっていると便利かをはっきりとイメージする。(前のページに例に挙げたようなイメージと比較してどうか)

(2) 検索項目入力画面の検索項目の分類はどのようなものがよいか。

(3) 何を検索項目とするか。ただし、検索項目に入れたものはNPLファイルまたはLSQファイルの段階で必ず正しい情報が入力されていることが必要である。検索する側からは検索項目は多い方がよいだろうが、NPLファイルのヘッダ部分や試料情報をすべて正確に記入したファイルを作成するのは大変であろう。データを提出する側からは、必ず正確な情報を入力しなければならない項目はあまり多くない方が、データの収集の面からも好ましいと思われる。従ってこの両者の折り合いをどこでつけるかというのが重要である。

(4) 機関名など、同じことを表すのに幾通りもの書き方がある項目をどのようにするか。

(5) サンプルコメントのように、どのような言葉でも入力される項目を検索できるようにするにはどのような方法を用いたらよいか。同じ意味に対して違う表現を用いている場合に、漏れこぼしなく検索するにはどのようにしたらよいか。

(6) IP接続以外でのデータベースアクセスについては、どのような接続方法でどのようなことができるのか勉強会を行って、ネットワーク環境の整備進行状況をにらみつつ検討したい。

5. 将来の課題

(1) デプスプロファイルのデータ

おそらく含まれるスペクトル1本1本を検索するというのではなく、まとまりとして利用すると考えられるので、上で述べたものとは異質のデータベースとなるであろう。したがって、サーバ内のMaster Indexの作り方自体が異なったものになるし、別のRDBMSを採用した方がよい場合もあるであろう。まずは、デプスプロファイルをデータベース化して何に利用するのか議論を行い、ある程度の利用イメージを作ることが必要である。次にデプスプロファイルのデータがいくつも共通で扱えるようになっていくことが、デプスプロファイルのデータベース化には必要である。これらをふまえた上で採用するRDBMS、Master Indexの作り方が決定されていくと思

われる。RDBMSのよい所は、Oracleとは別のRDBMS上にデプスデータベースを構築したとしても、クライアント側からはRDBMSの違いを感じることなく全く同等にどちらのデータベースへも自由にアクセスできることである。

(2) スペクトル形状による検索

特定の複数ピークが存在しているものや複数ピークの強度比がある一定範囲内にあるものなど、既にある程度イメージがあるように思われる。問題は、現在市販されているUNIX用RDBMSには、著者の知るところ、このようなパターン認識検索機能を持ったものがないことである。IRやNMRなどのスペクトルのパターン認識検索をUNIX上でやっている方があれば、是非表面分析研究会で報告し、JSA誌に解説記事を書いて下さい。コンピュータでのイメージ入力や指紋照合などパターン認識技術が用いられている分野は既に存在すると思われるので、パターン認識機能を持ったRDBMSが市販されるのもそう遠い将来ではないと期待している。

(3) 分析装置付随のコンピュータからのアクセス

データベースにアクセスするのはデータ解析に利用するためと思われるので、当然、分析装置に付随するコンピュータから直接データベースにアクセス出来るほうが望ましい。現状では、分析装置に付随するコンピュータから出力されるデータは形式が統一されていないので直接アクセスすることは不可能である。分析装置に付随するコンピュータからテキストファイル形式で出力できるようになったのもここ数年の変化であることを考えれば、ある程度ISOフォーマットが決定し認識されるようになれば、分析装置に付随するコンピュータからの出力がISOフォーマットになることも考えられる。そうなれば、検索プログラムを汎用性の高いUNIX上の言語で作成し、分析装置に付随するコンピュータから直接データベースにアクセスすることが可能になるであろう。

過渡的に、高頻度で利用する参照スペクトルをサーバから分析装置に付随するコンピュータにコピーしておきたいという場合には、現在ボランティアにより行われているフォーマット変換サポート作業と同様に、フォーマット逆変換サポートをボランティアにより行うことも考えられる。

最後に、UNIXデータベースやSQL言語について勉強したい、COMPROを使わずデータベースにアクセスしたいという方は、UNIXデータベースの入門書としては、UNIXリテラシー・シリーズVol. 15「UNIXデータベース」(箕田稔、津久井誠著、IBJ出版、2500円)など、SQL言語についてはOracleの解説書が多種出版されているので、これらを参考にしてください。