

УДК 543.442; 539.26; 546.06;

Язык запросов для рентгенофазовой идентификации многофазных поликристаллических материалов

И.С. Якимов*

Сибирский федеральный университет,
Россия 660025, Красноярск, пр.Красноярский рабочий, 95¹

Received 16.11.2009, received in revised form 7.12.2009, accepted 15.12.2009

Информационно-поисковая система рентгенофазовой идентификации (ИПС ФИ) вещественного фазового состава сложных многокомпонентных поликристаллических материалов в ходе Интернет-конкурса Search-Match Round Robin – 2002, проведенного комиссией CPD Международного союза кристаллографов, обеспечила точную идентификацию всех конкурсных материалов. Базисным элементом ИПС ФИ является проблемно-ориентированный операторный язык поисковых запросов, который обеспечивает высокую селективность фазовой идентификации и представляет альтернативу традиционным детерминированным алгоритмам и программам Search-Match.

Ключевые слова: рентгенофазовый анализ, фазовая идентификация, количественный рентгенофазовый анализ, метод ссылочных интенсивностей.

Введение

Фазовая идентификация (ФИ) или качественный рентгенофазовый анализ [1], является распространенным методом анализа вещественного фазового состава поликристаллических материалов и основана на компьютерном сопоставлении (Search/Match) рентгеновского дифракционного спектра (дифрактограммы) с эталонными линейчатыми дифракционными спектрами (штрих-спектрами) стандартов чистых фаз, накапливаемыми в базах данных (БД). В наиболее распространенной БД PDF ICDD в 2009 г. накоплено около 700 тыс. рентгенофазовых стандартов неорганических и органических веществ (<http://www.icdd.com>). При повышении сложности фазового состава или большом ко-

личестве фаз эффективность идентификации существенно падает, что выражается в отборе идентификационной программой большого числа спектрально близких фаз-претендентов. В результате ФИ сводится к трудоемкому визуальному сопоставлению эталонных спектров фаз-претендентов с дифрактограммой на экране монитора. Ограниченные возможности программ ФИ при идентификации дифрактограмм образцов сложного фазового состава показал конкурс Search-Match Round Robin – 2002 [2], проведенный комиссией CPD Международного союза кристаллографов.

Задача ФИ по своей природе является поисковой. Автором развит подход, основанный на использовании проблемно-ориентированного языка запросов и реа-

* Corresponding author E-mail address: I-S-Yakimov@yandex.ru

¹ © Siberian Federal University. All rights reserved

лизованной на его основе информационно-поисковой системе фазовой идентификации (ИПС ФИ), включающей интерпретатор запросов, графический интерфейс пользователя и нестандартный количественный анализ по мультирефлексному регуляризованному методу ссылочных интенсивностей (RRIR) [3]. ИПС ФИ обеспечивает автоматическое программирование скрытых от пользователя запросов на ФИ в результате оперирования мышью на графических окнах с дифрактограммой, эталонными спектрами, таблицей Менделеева и т.п., визуальный контроль спектров фаз-претендентов и построение из них модельного спектра дифрактограммы в качестве критерия достоверности ФИ [4]. В настоящей работе представлено описание языка запросов ФИ и его аналитических возможностей. Применение языка запросов позволяет пользователю реализовывать различные стратегии идентификации вместо использования программ Search/Match с детерминированными алгоритмами. Данный подход позволяет существенно сократить количество отбираемых фаз-претендентов и повысить эффективность анализа материалов со сложным составом, например, только ИПС ФИ дала возможность совершенно точно идентифицировать фазовый состав конкурсных заданий SMRR-2002 [5].

Язык запросов фазовой идентификации

Процесс фазовой идентификации рассматривается как последовательность идентификационных запросов, создающих выборки из базы данных эталонных спектров или из предыдущих выборок, в результате которого создается итоговая выборка, содержащая идентифицированные эталонные спектры фаз анализируемого материала.

Выборки сортируются по вероятностному спектральному критерию соответствия эталонных спектров с идентифицируемым спектром. Каждый запрос является интерпретируемой ИПС программой, состоящей из последовательности операторов языка ФИ. Программы-запросы позволяют описывать априори известную информацию о возможном фазовом составе в терминах задания допустимых комбинаций химического состава и формировать спектральные критерии соответствия эталонных и экспериментальных спектров для селективного поиска и отбора адекватных фазовых эталонов.

Для описания конструкций языка ФИ используется ряд мнемонических обозначений: «...» – скобки для содержательного описания элементов языка; [...] – скобки для необязательных языковых конструкций; {... | ...} – скобки для альтернативных языковых конструкций («или»). Другие мнемонические обозначения вводятся далее по тексту.

Формат запроса

IN = « Идентификатор входного объекта или операция над объектами»;

[«Последовательность операторов»]

OUT = « Идентификатор выходного объекта»;

где

«Объект» есть {«Спектр» | «Выборка номеров спектров»}.

«Спектр» есть {«Идентифицируемый» | «Эталонный» | «Модельный»} штрих-спектр в формате базы данных PDF ICDD.

Операторы отбора по хим. составу

{YES | NOT} = IF {ONE | ALL} {(F<L) | (F#<L) | (L<F) | (L#<F)};

LIST = «список {«химических элементов» | «фрагментов химических формул»}»;

Мнемонические обозначения: YES | NOT – отбор или отбраковка спектра из IN, если выполнено условие IF; F – хим. состав химической формулы фазы спектра из IN; L – хим. состав из списка LIST; «<» – знак «содержится»; «#<» – знак «не содержится»; ONE | ALL – кванторы «содержится хотя бы один» или «содержатся все».

Операторы отбора

по указанным линиям

$\{D2\Theta | d | \Delta H\} = \{ \Delta 2\Theta | \Delta d | \Delta H \} ;$
 $\{ Y E S | N O T \} = I F \{ O N E | A L L \}$
 $\{ (L < S) | (L \# < S) | (S < L) | (S \# < L) \}$
 $[FOR\{(I \geq I_{min})|(C * I \geq I_{min})\}] ;$

LIST = «список положения линий в шкале {2Θ | d | H}»;

где 2Θ – угол дифракции, d – межплоскостное расстояние, H=целое(10000/d);

Δ2Θ | Δd | ΔH – допустимая погрешность соответствия линий по 2Θ | d | H;

Мнемонические обозначения: S – очередной эталонный спектр из IN, сопоставляемый с идентифицируемым спектром IS; L – список позиций линий IS, указываемых в LIST для проверки соответствия им линий из S; FOR(I ≥ I_{min}) – проверка осуществляется только для линий S, имеющих интенсивность I ≥ I_{min}; FOR(C*I ≥ I_{min}) – проверка осуществляется только для линий S, имеющих интенсивность C*I ≥ I_{min}, где C – коэффициент приведения интенсивности линий спектра S к IS.

Операторы отбора

по спектральным критериям

$\{YES | NOT\} = IF (Rs \{ \geq | < \} Ro) [FOR \{(I \geq I_{min}) | (C * I \geq I_{min})\}] ;$

Мнемонические обозначения: Rs – идентификатор вероятностного критерия парного соответствия спектров, рассчитываемого для очередного спектра из IN и

спектра IS; Ro – пороговое значение критерия Rs.

Вероятностные критерии парного соответствия спектров рассчитываются на основе аппроксимации распределения интенсивности линий штрих-спектров в виде отношения интеграла перекрытия эталонного спектра S и идентифицируемого спектра IS к интегралу спектра S:

$$Rs = \int [Y(I^S(H_j), \Delta H, H) \cap Y(I^{IS}(H_j), \Delta H, H)] dH / \int [Y(I^S(H_j), \Delta H, H)] dH,$$

где Y(I(H_j), ΔH, H) – функция распределения интенсивности линий j с центром H_j, интенсивностью в максимуме I(H_j) и условной шириной 2ΔH, задающей погрешность совпадения эталонной и экспериментальной линий в целочисленной шкале обратных межплоскостных расстояний H = целое[10000/d]. При построении вероятностных критериев могут использоваться различные формы аппроксимации распределения интенсивности линий в задаваемом интервале соответствия 2ΔH, например нормальное, треугольное, прямоугольное, дельта-функцией. При этом могут быть применены различные весовые схемы учета вклада интенсивности линий в интеграл. Следует уточнить, что аппроксимируется не профиль линии, а условное распределение интенсивности, зависящее от ширины интервала 2ΔH и используемое только для вероятностной оценки степени соответствия спектров.

Свойства и правила взаимодействия

операторов

Важнейшим свойством индивидуальных операторов является их двойственность: эквивалентность действия при замене входящих в оператор операций на противоположные. Например, эквивалентны пары операторов:

YES = IF ONE (F < L) и NOT = IF ALL (F # < L);

YES = IF ALL (L<S) FOR (I>=50) и NOT = IF ONE (L#<S) FOR (I>=50);

YES = IF (Rs >=20) FOR (C*I>=10) и NOT = IF (Rs <20) FOR (C*I>=10).

Двойственность позволяет ввести в язык ФИ следующие правила для конструирования запросов из последовательностей операторов:

- 1) объекты, содержащиеся в IN, обрабатываются операторами запроса последовательно;
- 2) объект из IN отбирается в OUT тогда и только тогда, когда одновременно: не выполняется ни один из операторов NOT и выполняется хотя бы один из операторов YES;
- 3) отсутствие оператора YES эквивалентно его выполнению.

Критерий достоверности ФИ

Критерием достоверности ФИ является модельный спектр, интерпретирующий химический состав и все линии IS. Модельный спектр определяется как линейная комбинация из штрих-спектров адекватно идентифицированных рентгенофазовых эталонов БД PDF:

$$I^M(2\Theta_j) = \sum_i^m C_i \cdot I^S(2\Theta_{ij}), \quad j=1, \dots, n,$$

где m – количество фаз, n – количество линий, C_i – коэффициенты приведения спектров S_i к IS, $2\Theta_{ij}$ – позиции линий фазы S_i , $I^S(2\Theta_{ij})$ – интенсивности линий фазы S_i .

Коэффициенты приведения определяют и далее используются для нестандартного количественного фазового анализа по методу RRIR в соответствии с [3].

Обсуждение идентификационных возможностей языка запросов

Для идентификации образцов сложного фазового состава разработана методика «кластерной фазовой идентификации»,

основанная на использовании нескольких стратегий ФИ, разбивающих множество претендентов на группы (кластеры) так, что в каждом кластере присутствует только одна подлежащая идентификации истинная фаза (или ни одной, при неадекватном запросе). Базовыми являются стратегии «Поиск по сильной неидентифицированной линии дифрактограммы» и «Поиск по неидентифицированному химическому элементу образца». Соответственно, в первой стратегии в запрос, кроме операторов отбора по спектральным критериям, включаются операторы отбора по указанным линиям, а во второй – операторы отбора по указанным химическим элементам или фрагментам химических формул. В более сложных случаях применяется комбинация нескольких операторов отбора и отбраковки по линиям и элементам, например для проверки гипотез о принадлежности нескольких линий или (и) элементов к одной фазе.

Рассмотрим применение стратегий на примере идентификации на 2-м этапе конкурса SMRR-2002 одного из образцов (Sample 1) с указанным элементным составом (Fe, Ca, Al, Si, O, C, P, F, H), состоящим из четырех минеральных фаз (Siderite, Fluorapatite, Gormanite, Quartz), имеющих сильно искаженные дифрактограммы. Идентификация выполняется с помощью пяти запросов, изображенных в левой части рис. 1-5. В правой части рисунков приведен отобранный по запросу кластер фаз (истинная фаза подчеркнута) и изображен фрагмент дифрактограммы (излучение $\text{Cu}_{K\alpha}$) с модельным спектром, построенным из идентифицированных фаз. На рис. 1 приведен запрос на создание из БД генеральной выборки S1 по заданному элементному составу, включающей 4200 эталонных спектров фаз с указанным составом.

На рис. 2 отражен запрос на создание из S1 выборки S8, включающий операторы отбо-

```

IN=SS;    «БД»
YES=IF ALL(F<L);
LIST=H,C,O,F,Al,Si,P,Ca,Fe;
OUT=S1;  «выходная
          выборка»

```

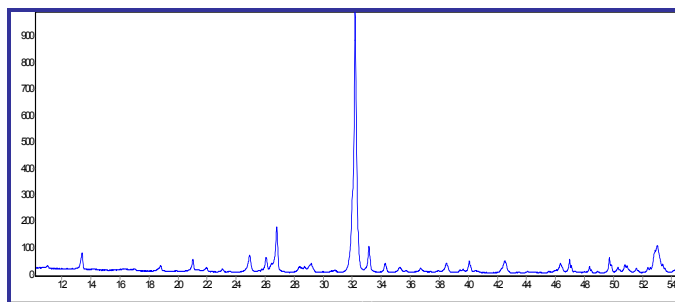


Рис. 1. Запрос на создание генеральной выборки по элементному составу и дифрактограмма идентифицируемого образца Sample 1

```

IN=S1;
DH=40;
NOT=IF(RHC<15);
NOT=IF
ONE(L#<S)FOR(I>=50);
LIST=3601; «2θ~32.2°»
NOT=IF ALL(L#<F);
LIST=Fe;
YES=IF ALL(F<L);
LIST=Fe,H,C,O,F,Al,Si,P,Ca;
OUT=S8;

```

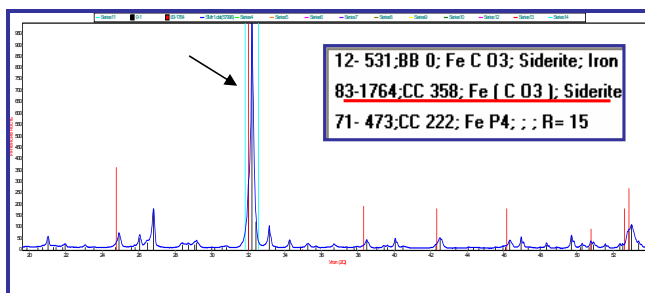


Рис. 2. Запрос на создание выборки по сильнейшей линии и основному элементу (Fe), справа в рамке – состав выборки по запросу, красные штрихи на дифрактограмме – идентифицированный спектр сидерита (№83-1764)

```

IN=S1;
DH=30;
NOT=IF(RN<100)FOR(I>=50);
NOT=IF(RN<50)FOR(I>=25);
NOT=IF(RHC<15);
NOT=IF
ONE(L#<S) FOR(I>=34);
LIST=3009; «2θ~26.7°»
OUT=S8;

```

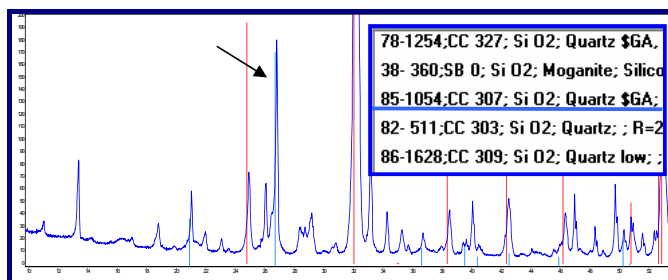


Рис. 3. Запрос на создание выборки по сильнейшей неидентифицированной линии, справа в рамке – состав выборки по запросу, синие штрихи на фрагменте дифрактограммы – идентифицированный спектр кварца (№85-1054)

```

IN=S1;
DH=30;
NOT=IF(RN<100)FOR(I>=50);
NOT=IF(RN<50)FOR(I>=25);
NOT=IF(RHC<15);
NOT=IF ONE(L#<S)
FOR(I>=70);
LIST=3704; «2θ~33.1°»
NOT=IF ALL(L#<F);
LIST=P;
YES=IF ALL(F<L);
LIST=P,H,C,O,F,Al,Si,Ca,Fe;
OUT=S8;

```

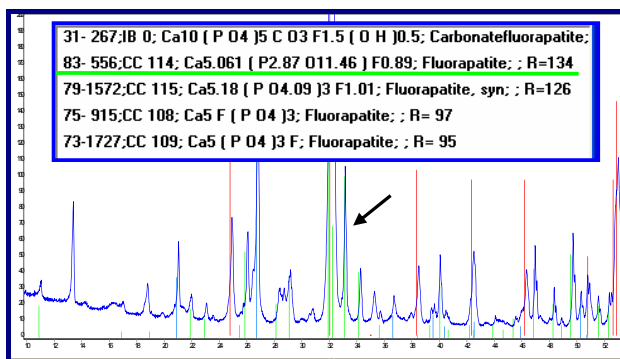


Рис. 4. Запрос на создание выборки по сильнейшей неидентифицированной линии и неидентифицированному элементу (P), в рамке – состав выборки, зеленые штрихи на фрагменте дифрактограммы – идентифицированный спектр фторапатита (№83-556)

```

IN=S1;
TVTYPE=5;
DH=40;
NOT=IF(RN<100)FOR(I>=50);
NOT=IF(RN<50)FOR(I>=25);
NOT=IF(RHC<15);
NOT=IF ONE(L#<S)
FOR(I>=100);
LIST=1511; «2θ~13.3°»
NOT=IF ONE(L#<S)
FOR(I>=15);
LIST=2119; «2θ~18.8°»
OUT=S8;

```

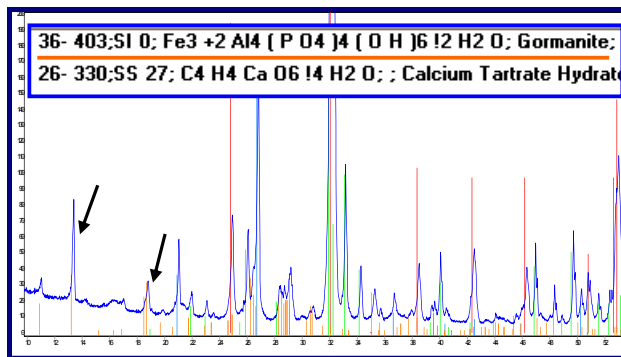


Рис. 5. Запрос на создание выборки по двум сильнейшим неидентифицированным линиям, справа в рамке – состав выборки по запросу, коричневые штрихи на дифрактограмме – идентифицированный спектр горманита (№36-403)

Таблица 1. Результат КРФА образца Sample1 SMRR-2002

N PDF	Formula	Mineral Name	% масс.
83-1764	Fe (C O ₃)	Siderite	46,6
85-1054	Si O ₂	Quartz	13,3
83- 556	Ca _{5,061} (P _{2,87} O _{11,46}) F _{0,89}	Fluorapatite	29,7
36- 403	Fe ₃ ² Al ₄ (PO ₄) ₄ (OH) ₆ 2H ₂ O	Gormanite	10,2

ра по спектральным критериям соответствия, сильнейшей линии и основному элементу образца (Fe). В результате запроса создана выборка S8 из трех эталонных спектров фаз-претендентов, в которой первые два являются спектрами сидерита. Визуальное сравнение этих спектров с дифрактограммой идентифицировало сидерит $\text{Fe}(\text{CO}_3)$. В то же время, при исключении из данного запроса оператора отбора по сильнейшей линии сидерит находится в выборке на 11-м месте, при исключении оператора отбора по основному элементу – на 17-м, а при исключении обоих операторов (т.е. при поиске только по спектральному критерию соответствия) – на 24-м.

Сложность идентификации сидерита в данном образце обусловлена дефектной структурой и сильной текстурой вдоль направления (112), отразившихся в сдвиге его линий на дифрактограмме на величину $\Delta 2\Theta = 0.15\text{--}0.20^\circ$ и аномально высокой интенсивности сильнейшей линии относительно эталонных спектров. Использование в запросе вышеописанных операторов приводит к его однозначной идентификации.

На рис. 3 изображен запрос на создание выборки, включающий операторы отбора по спектральным критериям соответствия и сильнейшей неидентифицированной линии. В результате запроса создана выборка S8 из пяти эталонных спектров фаз-претендентов, в которой отобран кластер из эталонных спектров кварца.

На рис. 4 приведен запрос на создание из S1 выборки S8, включающий операторы отбора по спектральным критериям соответствия, сильнейшей неидентифицированной линии и неидентифицированному элементу (P). В результате запроса создана выборка S8 из пяти эталонных спектров фаз-претендентов, в которой отобран кластер из эталонных спектров фторapatита.

На рис. 5 приведен запрос на создание из S1 выборки S8, включающий операторы отбора по спектральным критериям соответствия и двум оставшимся неидентифицированным линиям. В результате запроса создана выборка S8 из пяти эталонных спектров фаз-претендентов, в которой отобран кластер с эталонным спектром горманита на 1-м месте.

В табл. 1 представлены результат КРФА данного образца по методу RRIR. Язык запросов ФИ основан на элементах математической теории исчисления предикатов [6], в частности, на использовании кванторов «для всех» (ALL) и «существует» (ONE). Это обеспечивает его более высокие поисковые возможности по сравнению с традиционно применяемой в программах ФИ булевой логикой и создает возможность реализации селективных стратегий ФИ. Интерпретатор языка запросов, средства их графического автопрограммирования и ИПС ФИ разработаны в среде визуального программирования Дельфи. Аналоги языка запросов ФИ могут быть реализованы и на реляционно полных языках манипулирования базами данных с процедурными расширениями, типа SQL/PSM.

Заключение

Стратегии поиска по выделенной линии и/или элементному составу приводят к кластеризации множества возможных фаз-претендентов. В кластер при подходящих значениях поисковых критериев попадает группа из нескольких спектрально и, как правило, кристаллохимически подобных фаз, среди которых находится одна истинная фаза, имеющая более высокое значение рейтингового критерия по сравнению с большинством фаз-претендентов. Кроме того, истинная фаза легко определяется в кластере визуально – по соответствию ее эталонного спектра с дифрактограммой образца.

Как следствие, повышается достоверность ФИ, экспрессность анализа (~1-5 мин/фаза) и устойчивость МНК при синтезе модельного спектра образца из идентифицированных эталонных спектров. Модельный спектр служит критерием достоверности ФИ и используется для автоматического КРФА по мультирефлекскому регуляризованному методу ссылочных интенсивностей. Приме-

нение языка запросов и основанной на нем методики кластеризации, реализуемой интерактивно с помощью ИПС ФИ, существенно упрощает процедуру фазовой идентификации и КРФА материалов со сложным фазовым составом, а главное, создает у пользователя уверенность в том, что истинные фазы не были опущены в рассмотрении и действительно идентифицированы.

Список литературы

1. Васильев Е.К., Нахмансон М.С. Качественный рентгенофазовый анализ. – Новосибирск: Наука, 1986. – 192с.
2. J-M Le Meins, L.M.D. Granswick, A. Le Beil. Results and conclusions of the internet based «Search/match round robin 2002» // Powder Diffraction – June 2003 – Volume 18, Issue 2, pp. 106-113.
3. Якимов И.С., Дубинин П.С., Пиксина О.Е. Регуляризация метода ссылочных интенсивностей для количественного рентгенофазового анализа поликристаллов // Журнал Сибирского федерального университета. Серия «Химия». – 2009. – Т. 2, №1. – С.71-80.
4. Yakimov I. S. System of X-Ray Identification of Substantially Multiphase Materials // Inorganic Materials. – 2008, -Vol. 44, -No. 14, -p. 54–58.
5. Search-Match Round Robin – 2002 // <http://sdpd.univ-lemans.fr/smrr/results/>
6. Мендельсон Э. Введение в математическую логику. – М.: Наука, 1984. – 300 с.

Query Language for X-Ray Powder Diffraction Phase Identification

Igor S. Yakimov
Siberian Federal University,
95 Krasnoyarskii Rabochii, Krasnoyarsk, 660025 Russia

The information storage and retrieval system of X-ray powder diffraction phase identification (Retrieve) identified all Search-Match Round Robin-2002 samples exactly. A base element of the system is a problem-oriented operator language of search queries for phase identification. The query language is described in detail and its identification possibilities are discussed.

Keywords: X-ray powder diffraction, qualitative phase analysis, XRD phase identification, quantitative phase analysis, reference intensity ratio method.
