Information-Calculation System "WWW-MINCRYST" – Web-Toolkit for Working with Mineral Structures and its Integration into the Space of Geological Knowledge

Dmitry Varlamov [1][0000-0002-8474-7851]

[1] Institute of Experimental Mineralogy RAS, 4 Ossypyana, Chernogolovka, Russia, 142432
dima@iem.ac.ru

Abstract. In article are considered history of creation and development (more than 23 years) of the WWW-MINCRYST information-calculation system (ICS) intended for working with crystalline structures of minerals and their synthetic analogues. The structure and implementation methods of system, structure and functions of principal components are described: the database (more than 11,100 records for more than 4,600 unique phases); a complex of intellectual tools of search and choice of information, including original ones; tools for multimedia presentation of information (multifunctional interactive structures and spectra); the possibilities of processing of spectral information, the means of forming cross-references with external web-resources and the use of auxiliary information. The main problems in the development of ICS are shown, related to the trends in the development of the main WWW browsers, the change of web technologies and security problems. Particular attention is paid to the integration of ICS with external resources and vice versa. The basic problems in the development of ICS and inter-resource exchanges of information, security and the formation of mixed content are considered. The ICS is hosted at http://mincryst.iem.ac.ru.

Keywords: Internet-oriented systems, minerals, crystal chemistry, databases, search criteria, interactive multimedia applets

1 Introduction

Information-scientific Internet resources (including those on Earth sciences) can be subdivided [1] by the type of content into the following basic categories:

- descriptive (collections of articles, monographs, lectures, descriptions, including those in the Wiki-style format);
- events (event monitoring [earthquakes, eruptions etc., current news, conferences);
- discussion (forums, blogs, discussions, FAQ, social networks);
- reference (databases, catalogs, libraries);

- interactive multimedia pages (modeling of objects and processes, specialized calculations, GIS, demonstration programs)
- software storages, computing resources.

The presentation of databases and interactive resources on the Internet still causes certain technical and conceptual difficulties. The main problems include heterogeneous (often incomparable) structures of heterogeneous data, the lack of standards for the presentation of specialized information, the "diversity" of interfaces to existing databases, differences in the tasks of search agents and automatic compilers of information and forms of its presentation.

The authors, using the example of two information-computing (ICS) web-oriented systems, combining descriptive, reference and multimedia resources, show options for creating, maintaining and updating convenient (in our opinion) user-friendly interactive systems for presenting specialized data.

2 ICS «WWW-MINCRYST»

At the end of 1997, on the basis of the first tools of developing interactive Internet resources that arose, the "WWW-MINCRYST" database was implemented, which became one of the first interactive Internet resources in the field of Earth sciences in Russia and in the world [2]. The initiator of the work was the head of the group of X-ray analysis in IEM RAS outstanding crystallographer A.V. Chichagov (alas, who died in 2010). The purpose of the resource was to provide interactive Internet access for users to large arrays of mineralogical and crystallochemical information. The ICS "WWW-MINCRYST" factual basis was provided by the literature data accumulated since 1985 on the crystal structures of minerals, compiled into a local PC database (based on literary sources and author's analytical data - about 3,500 objects at the start of the project).

Further, the ICS "WWW-MINCRYST" (http://mincryst.iem.ac.ru) was created with tools for interactive data processing and analysis. The basics of the ideology and technology of ICS are described here [3–5]. ICS "WWW-MINCRYST" was originally based on the use of Internet technologies and has become one of the pioneer interactive scientific Internet resources both in Russia and in the world (in the field of Earth sciences). The first fully operational version of the WWW interface was created under the RFBR grant No. 96-07-89162 and was presented to users in December 1997, i.e. ICS has been functioning, developing and is actively demanded by users for almost 23 years.

The purpose of creating the ICS "WWW-MINCRYST" is to provide as wide a range of users as possible in all areas of science operating with crystalline matter (geology, geochemistry, crystallography, solid state physics, surface physics, materials science, etc.) with reliable and relevant up-to-date information in the field of structural crystal chemistry of minerals, their synthetic analogues and elements, to develop methods of processing and visualization of crystal structure data.

During the ICS-project, "WWW-MINCRYST" increased its information fund by almost 3 times (by more than 7500 objects) and rightfully entered the front ranks of X-ray structural and crystal-chemical databases related to the study of mineral matter, both

according to users' estimates and according to compilers of industry catalogs of information links.

3 Basic System Components and Technical Implementation

The main component of the ICS “WWW-MINCRYST” is the database itself or the Information Fund (as of September 2020 – more than 11,100 information objects for almost 4,600 unique phases, including 4,400 natural minerals and 200 synthetic phases that have no natural analogues yet). The information object (card) consists of 4 base tables with 60 fixed and up to 130 additional fields (the number of the latter varies depending on the complexity of the structure) and 4 indexed binary files. A hierarchical index system is supported. The total database size is about 600 MB.

The Information Fund contains information on most mineral species (about 4,400 out of almost 5,900 officially recognized by the International Mineralogy Association as of March 2020 – http://www.ima-mineralogy.org/Minlist.htm), the crystal structures of which have been decrypted to date. In addition to natural mineral species, the database contains synthetic minerals – their structural analogues that differ in composition (for example, with the replacement of one of the cations), and inorganic compounds (silicates, phosphates, borates, etc.) that are close in properties to natural substances. The information fund contains structural work data from more than 130 foreign and domestic journals for the period from the 1930s to the present. The annual growth of new or newly redefined/refined crystalline structures of minerals and their analogues is significant enough to require constant updating of the information fund (on average, up to 350–450 new structures per year + 150–200 structures subjected to revision and changes). In recent years, the main emphasis has been placed on new minerals, actively replenished including by Russian researchers.

The user interface of ICS “WWW-MINCRIST” is a multilevel information bilingual (ENG/RUS interface and mineral names) system. The interface includes: (a) complex search interfaces using mineral names, chemical composition, crystal structure parameters, literature references and supporting information as search criteria (in various combinations), and combinations of search features are possible; (b) modules for multimedia presentation of information (interactive polyhedral and spherical structures, linear and continuous spectra – through Java applets “WWW-CrystPic” and “WWW-MixiPol”; (c) classification schemes of several authors; (d) modules for organizing relationships with external information resources through a system of dynamically generated links; (e) WWW-oriented developer tools (including a system for importing, verifying and editing data, as well as archiving and backup).

Technologically, the ICS “WWW-MINCRIST” as a web service is implemented on the "classic" Linux-Apache-MySQL-PHP (LAMP) bundle using JavaScript and (for interactive applets) Java (based on Sun/Oracle JDK). The ICS is located on the IEM RAS database server, access is not limited. At the launch of the ICS project, “WWW-MINCRIST” became in 1998 (implemented on the Digital Unix OS) one of the first Internet-oriented databases in Russia to use this technology, and so far it has not lost its

technicality on the server side. In 2004, the ICS was implemented under Linux environment; in 2015, the database was migrated from the proprietary MySQL DBMS to the freely distributed MariaDB DBMS (https://mariadb.org) to maintain the free status of the ICS. Currently, the ICS is hosted on a server with the following parameters – Intel Core i5-2400@3.10GHz x 4 cores, RAM 24 Gb, disk array SSD 128 Gb + 2x2 Tb WD RE 2000, GPU Nvidia Kepler GTX 770, software – Apache HTTPD server 2.4.6, PHP script language 5.4.16, MariaDB database 5.5.40.

Journal articles are the main source of information for ICS. The extracted crystal structure information is placed (in accordance with a special recording format) in an ASCII file with subsequent software examination (conducted by expert authors on local PCs) based on the calculation of interatomic distances and other crystal structure characteristics and, in the case of a positive solution, is imported into the ICS by special tools via the web-interface both in the form of single records and packet records (up to several hundred) with incoming data control, which allows continuous updating of the information fund. It is also possible to edit records in on-line mode, delete them, replace service records files, archive and restore the database via the web interface.

The information record for an individual crystalline structure contains information on the name (in accordance with the classification of the International Mineralogy Association or IUPAC inorganic naming guidelines), chemical composition, symmetry, unit cell parameters, atomic position coordinates with isotropic temperature factors and populations, information about interplanar distances, HKL indices and intensities of the strongest reflections of the X-ray diffraction pattern of the polycrystal phase, as well as links to the corresponding general publications on the interpretation or refinement of the crystal structure. The record can be specified by useful properties, chemical composition and structure, P–T conditions of synthesis, and belonging to conventional mineral groups. Each record contains "monocrystal" and "polycrystal" characteristics of the crystalline phase. Minerals are classified according to the taxons of structural chemical classifications belonging to A.A. Godovikov, M. Chiriotti, G.B. Bokiy. For 2500 phases, express estimates of the potential energy of the crystal lattice were made. At present, a system is being developed for calculating the element contents in accordance with the ideal and real mineral formulas.

The ICS implements a system of phase search according to individual sets or a set of criteria: the name of the mineral (full or partial, according to the specification, as part of commonly used groups, alphabetical lists), chemical or elemental composition in various combinations (presence/absence of elements, their combinations, stable chemical groups, chemical classes of substances), crystal structural characteristics (symmetry, space groups, unit cell parameters, interplanar spacing d(hkl)), which in combination with chemical (elemental) composition enables direct interactive qualitative X-ray phase analysis based on the measurement results. A search has been added to the search system for the classification parameters of several crystal-chemical and structural classifications (Godovikov, Chiriotti, Bokiy). A search system by these parameters provides a search (and grouping) of minerals according to the indicated lower and intermediate taxon of classifications. According to a number of search parameters, the ICS “WWW-MINCRYST” still has no analogues among mineralogical-crystallographic databases.

Based on the results of the search, the user receives a link to the list of phases that meet the search criteria, specifying the record ID, phase name, formula and spatial group, with the possibility of subsequent transition through the hyperlink. In case of a single successful response, the user is immediately transferred to the found object if a group of objects is found - a selection list is offered.

In 2017–2019, important additions were made to the chemical composition search system for searching for stable combinations of elements (such as anions/cations: SO_4, CO_3; neutral groups – H_2O, NH_3; structural silicate motifs – SiO_3, Si_2O_7, Si_3O_8, etc. .), the possibility of searching by impurity isomorphic atoms that are not reflected in the idealized formulas of minerals (which is important for scattered elements); Search schemes based on the mutual absence-presence criteria of elements have been improved. The transition to the unified utf-8 encoding has been completed (for the internal presentation of textual information in the database and for issuing to the user), which simplifies working with modern browsers and allows using of diacritics in the spelling of the original mineral names. In the “WWW-Crystpic” module, a number of inaccuracies and remarks on the work with the reproduction of superstructural elements have been corrected, and the possibilities of working with large structures (in terms of the number of atoms, up to 200, for example, in complex structures of zeolites) have been expanded.

The “map” of the found object contains the basic data (both static from the database and dynamically calculated) – name, formula, cell parameters, various crystallographic data (number of reflexes, x-ray densities, absorption coefficients); CPDS card – 20 maximum reflexes, their *hkl* position and intensity; basic atomic positions and their settlements; complete information and calculation card (X-ray densities, all reflexes, etc.); calculated lattice energies; belonging to various classification schemes; data sources; automatically generated links to external resources and search queries.

For all records via the WWW-Crystal applet (so far – Java-3D using OpenGL libraries), dynamically created interactive images of crystal structure models in ball-spheres and in polyhedral projections (up to 138 positions and up to 1500 (!) atoms per structure) are available. The program allows you to to make all kinds of manipulations with the structure model, including scaling, continuous and / or automatic discrete rotation around the "screen" axes X, Y, Z; orientation along crystallographic axes, hkl-fragmentation of the structure (into hkl-oriented fragments of thickness d (hkl)), the growth of unit cells along any selected directions for the formation of "superstructures" and motifs, as well as a direct "manual" and automated for small polyhedra (tetrahedra and octahedra) calculation of any interatomic distances and angles (flat and solid) in the structure. The program depicts any polyhedra, including “defective” ones with unusually small (“bad”) interatomic distances. Examples of models of crystal structures (up to 600 atoms) are shown in Fig. 1

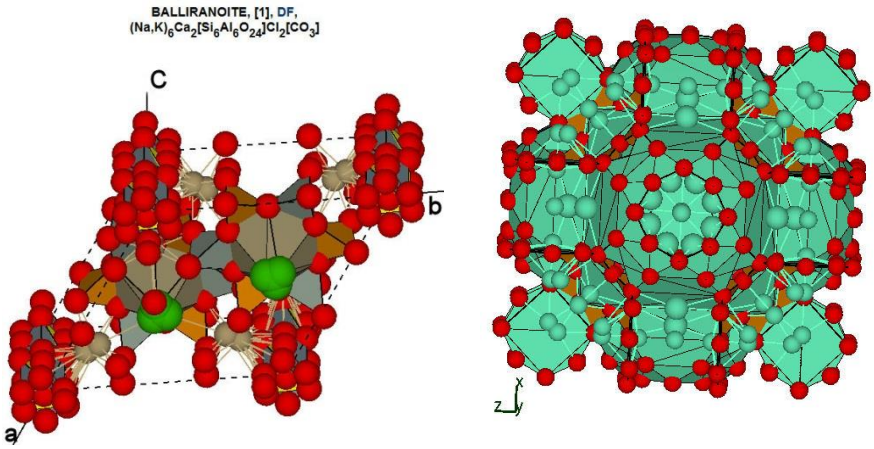

Fig. 1. Examples of interactive models of crystal structure (mixed representation in balls and polyhedral)

Currently, there are problems with working with external OpenGL libraries (on which the applet is implemented) and paranoid Java 8 security settings, which are solved by user intervention. However, because of this, porting the applet to new software platforms such as HTML5 or WebGL embedded tools is now imperative.

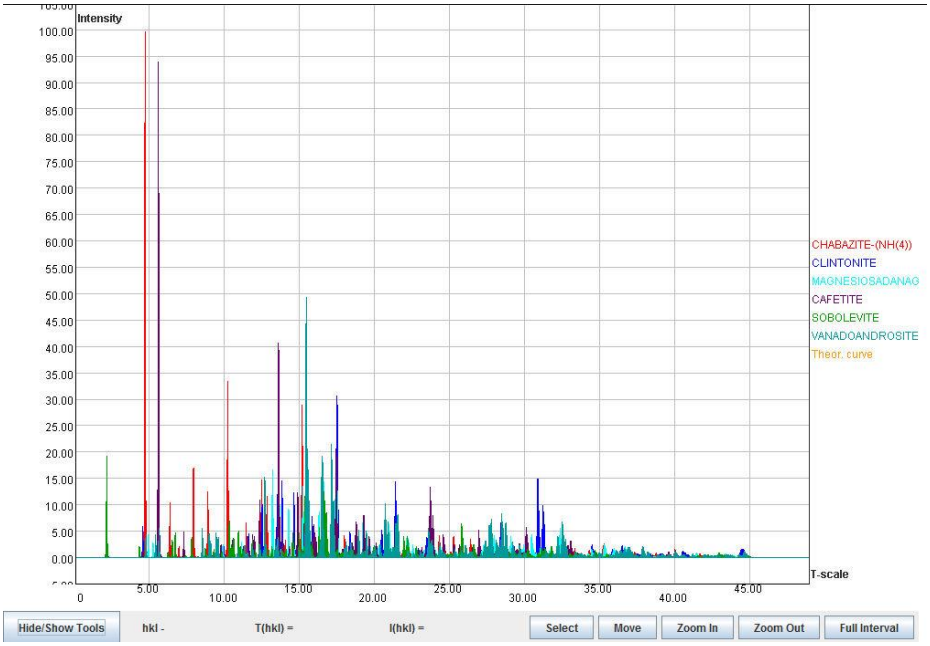

Fig. 2. An example of a combination of calculated spectral profiles for polycrystal x-ray spectrum (mixture of 6 phases)

The “WWW-Mixipol” applet module is designed to graphically represent the full calculated spectral profiles of polycrystal X-ray diffraction patterns with spectral manipulation capabilities for different radiation sources and different types of spectral scales. The module is also able to form X-ray diffraction patterns of phase mixtures (up to 6 phases simultaneously), with the possibility of varying the relative contents of the mixture components. An example of the calculated spectra of a mixture of six phases is shown in Fig. 2.

Both for structures and spectra of minerals, “simplified” presentation options are provided in the form of interactive traditional spherical structures and line spectra.

One of the first among the scientific Internet-oriented databases for the ICS WWW-MINCRYST a system of dynamically generated cross-web links was developed to link objects with records for specific minerals in the leading mineralogical databases on the Internet. The system for generating dynamic hyperlinks to external information resources (to mineralogical databases and search engines) allows you to "transparently" connect external data arrays to the user using the "general" query method [1]. In this case, the user immediately gets access to information on the object of interest to him, bypassing the stages of searching or viewing the entire external database. In addition, this mechanism implements feedback, allowing these databases to refer to **our** information objects in the same way, which sharply increases the demand for ICS by external users.

The availability in the ICS WWW-MINCRYST of data on 11 100 crystalline structures, an integrated computational complex, and developed visualization tools made it possible to use ICS in the development of non-traditional scientific approaches to the interpretation and presentation of certain crystalline structures, in addition to the capabilities oriented to searching and providing information. Processing of the array of accumulated data by complex SQL queries made it possible to obtain very non-trivial statistical samples, for example, according to the distribution in the nature of minerals among crystal symmetry groups. “WWW Xraypol” tool revealed the possibility of flexible use of polyhedra in traditional structures and almost automatically formed various variants of structural models of minerals. Due to the flexible use of polyhedra, which allows one to form various variants of structural models of minerals, for some minerals one can loosely bind to the traditional cationic-anionic image, and form structures based on any atoms that make up its composition.

The demand for ICS “WWW-MINCRYST” is well confirmed by the statistics of requests (in average 2015–2019 per year – 6–9 million successful single requests, more than 70–250 GB of downloaded information, 36,000 70,000 unique client sites, see section “Statistic” on the site), as well as a large number of reviews, descriptions and external links to WWW-MINCRYST (see section "Links").

4 Conclusions

The presented ICS "WWW-MINCRYST" is generally accessible, user-friendly interfaces to large arrays of crystal-chemical and mineralogical-petrological information with advanced search, presentation and processing tools and can serve as powerful tools

for all researchers in mineralogy, crystallography, petrology, solid state physics, materials science and other related fields of science.

Work on the “WWW-MINCRYST” and “WWW-Sapphirine” during 1997–2017 was supported by 7 RFBR grants.

References

1. Plechov, P.Iu., Varlamov, D.A., Trusov, S.V.: Tipy informatsionnykh potokov v oblasti nauk o Zemle. In: “Nauchnyi servis v seti Internet”, izd-vo MGU, Russia, 142–144 (2002).
2. Chichagov, A.V., Varlamov, D.A., Dilanyan, R.A. et al.: Crystallography Reports, 46(5), 876–879 (2001). doi 10.1134/1.1405882
3. Chichagov, A.V., Varlamov, D.A., Ershov, E.V., Dokina, T.N., Drozhzhina, N.A., Samokhvalova, O.L.: Zapiski RMO, 136(3), 135–141 (2007).
4. Varlamov, D.A., Dokina, T.N., Drozhzhina, N.A., Samokhvalova, O.L.: Herald of South-Urals State University, Ser. «Vychislitelnaia matematika i informatika», 2(1), 26–32 (2013). doi:10.14529/cmse130103
5. Varlamov, D.A., Podlesskii, K.K., Dokina, T.N., Drozhzhina, N.A., Samokhvalova, O.L.: Multimedia Web-based databases in Earth sciences (crystal chemistry, mineralogy, petrology). Nauchnyi servis v seti Internet: trudy XXI Vserossiiskoi nauchnoi konferentsii (23–28 sentiabria 2019 g., g. Novorossiisk). M.: IPM im. M.V. Keldysha, 2019. s. 198–209. doi:10.20948/abrau-2019-89.