

Raman Data Search and Storage (RDSS): a Java-based program for displaying and examining Raman spectra

Andrei Ionuț Apopei¹, Andrei Buzatu¹

¹ "Alexandru Ioan Cuza" University of Iași, Department of Geology, 20A Carol I Blv, 700505 Iași, Romania

Abstract

The ability to display and inspect Raman spectra quickly and efficiently is a central part of the data analysis process in Raman spectroscopy. In this manuscript, we present Raman Data Search and Storage (RDSS), a Java-based program that provides a user-friendly interface for the display and examination of Raman spectra. The RDSS allows users to easily visualize and analyze their data, facilitating the identification of characteristic peaks and patterns. Additionally, RDSS offers advanced search capabilities, enabling users to efficiently search and retrieve specific spectra based on various parameters such as peak intensity or wavelength range. Overall, the RDSS program is a valuable tool for researchers in the field of Raman spectroscopy, enhancing efficiency and aiding in the interpretation of complex spectra. With its intuitive interface and powerful search features, RDSS streamlines the process of data exploration and comparison, saving researchers valuable time and effort. Whether studying minerals, natural pigments, or other geological samples, RDSS software provides a comprehensive solution for analyzing Raman spectra with precision and ease.

Keywords: Raman spectroscopy, Raman data search, minerals, database, Java, spectral analysis

1. Introduction

Raman spectroscopy is quickly becoming one of the preferred analytical techniques, being used across a wide range of application areas, such as mineralogy (Nasdala et al., 2004), gemology (Bersani and Lottici, 2010), planetary geology (Sharma et al., 2003), archaeology (Vandenabeele, 2004), conservation (Vandenabeele et al., 2007), medicine (diagnostics, in vivo) (Hanlon et

^{© 2024} Ed. Univ. "Alexandru Ioan Cuza" Iași. All rights reserved Corresponding author's e-mail: andrei.apopei@uaic.ro

al., 2000), biomedicine (Popp et al., 2011), pharmaceutics (Vankeirsbilck et al., 2002) or forensics (Virkler and Lednev, 2010).

Given the fact that it is very fast, nondestructive, non-contact, and accurate for in-situ detection and characterization, this technique has proven extremely useful in various fields of materials science. In geosciences, the Raman technique is widely used due to its speed and the fact that it allows a first impression of the compounds present in the rock sample without the use of complex and time-consuming chemical analysis or other analytical techniques. For example, in gemology, this technique is used to discriminate between a genuine gemstone and a fake one. Additionally, Raman spectroscopy can help find plastics and epoxies that got into the samplemounting process and caused contamination (Lowry et al., 2009).

As Raman spectroscopy becomes more useful in many areas of materials science, more and more Raman spectra for mineral species are becoming available. Access to certain online databases¹ is, fortunately, free, while more extended databases with automatic search engines are commercially available, generally at very high prices. Given that, through an

¹ RRUFF project, *http://rruff.info*;

Spectroscopy, http://rdrs.uaic.ro;

online database, it is not possible to carry out the identification of spectra, at the moment, few programs cover this aspect of identification using the whole unknown Raman spectrum or Raman bands as identifiers and comparing them to the data in their own database.

Peak Finder² from the ColoRaman project (Burrafato et al., 2004) offers a search function that uses the wavenumber positions of the Raman bands through an input method called "seek." Unfortunately, this application has no GUI interface and can be run only on the Windows platform. CrystalSleuth (Laetsch and Downs, 2006) is a Windows-based piece of software capable of manipulating Raman spectra and, also, a great tool in the search for unknown spectra using similarity search algorithms that reference peak positions against spectra from the online RRUFF project (Downs, 2006). The manipulation of the spectra consists of removing background noise and cosmic ray events. The program also offers the possibility of comparisons between more Raman spectra in stacking mode. Another application available on the internet, called Raman Processing³, developed within the MATLAB environment, offers a wide range of features

http://cares.wayne.edu/rp/index.html

All accessed on September 2013

RASMIN, Raman Spectra Database of Minerals and Inorganic Materials,

http://riodb.ibase.aist.go.jp/rasmin/E_index.html; Minerals Raman Database,

http://www.fis.unipr.it/phevix/ramandb.php; RDRS, Romanian Database of Raman

Handbook of Minerals Raman Spectra,

http://www.ens-lyon.fr/LST/Raman/index.php; ColoRaman project, http://web.ct.infn.it/~archeo/;

Raman Spectroscopic Library of Natural and Synthetic Pigments,

http://www.chem.ucl.ac.uk/resources/raman/; IBeA (Ikerkuntza eta Berrikuntza Analitikoa), http://www.ehu.es/udps/database/index1.html; WURM project, http://www.wurm.info/

² ColoRaman project with Peak Finder utility, http://web.ct.infn.it/~archeo/

³ Raman Processing,

related to pre-processing, analysis, or even classification of the Raman spectra. Apart from these great options, however, this program does not come with an alreadydeveloped database of Raman references. A lot of research papers (Bakker Schut et al., 2002; Castanys et al., 2006; Ishikawa and Gulick, 2013; Perez-Puevo et al., 2004; Reisner et al., 2011; Salimbeni et al., 2005; Soneira et al., 2002; Vandenabeele, 2011) have previously dealt with the mathematical or statistical methods used to figure out Raman spectra. Such studies can help research groups get quick, correct, and reliable answers. Many of the algorithms are similar, being related to artificial intelligence, principal component analysis, cluster analysis, and correlation coefficients.

The main problem after acquiring the Raman spectra in the laboratory or in the field lies in the identification of the compound with high speed and precision. This is particularly true in the case of geological sites where exploration may be difficult or dangerous (e.g., remote exploration of the surface of Mars (Wang, 2003) or a deep-sea hydrothermal vent on Earth (White et al., 2006)). The process becomes even more complex when the Raman spectrum corresponds to a mixture of mineral species. There are, then, a wide range of associated issues: the technique employed must be usable without limitations, portable, user-friendly, and automated as much as possible. The search for similarities between a raw Raman spectrum and a large database may be a tedious task, dependent upon user experience. Typically, there are two ways to carry out

the identification process: (i) one consists of comparing the whole unknown (raw) Raman spectrum with the known standard spectra; (ii) the second strategy is based upon comparing the wavenumber positions of the Raman bands of the raw spectrum with those from the reference library. Problems such as noise or fluorescence are very common, and they render the analysis difficult and the search procedure uncertain.

This paper presents a simple, lightweight, and easy-to-access program for fast visualization and identification of unknown inorganic compounds. The program utilizes a reference library to compare Raman bands, ensuring accurate identification. This alternative solution offers a user-friendly interface and efficient search capabilities, making it a valuable tool for researchers and professionals in the field. Additionally, RDSS provides a cost-effective option compared to expensive commercial applications while maintaining a high level of accuracy and reliability. The Raman spectroscopy area offers unique challenges, and this software aims to address those challenges by providing researchers with the tools they need to analyze plain spectroscopic data.

2. Raman Data Search and Storage overview

The Raman Data Search and Storage (RDSS) software is written entirely in the Java language and uses the principles of the Java Virtual Machine (JVM) introduced by Sun Micro System (Lindholm et al., 2012). Java has the property known as "*write once, run anywhere (WORA)*," with the following characteristics: it is a simple, object-oriented, distributed, and interpreted programming language that offers high security, portability, and performance, as well as support for multi-threading. For example, a well-known Java-based program in the field of geosciences is *Jmol/JSmol* software (Hanson et al., 2013).

The *RDSS* application has a wellorganized and optimized layout for visual inspection of the specific Raman spectrum, and it lets you do certain things at the same time (Fig. 1). In terms of interface functions, the data are visualized by calling a specific dataset from the server (the MySQL database) and displaying it as a spectrum graph (Fig. 1–a). The Raman spectra can be manipulated using the mouse in combination with the keyboard.

The core of the graphic user interface consists of two tables that are populated with minerals and logically structured into columns (Fig. 1-b). A table data filter is applied to both tables. The header of each table has fields that allow the user to search by column of interest, for example, by selecting a specific mineral class. The search functions enable users to find a specific peak value. Using the asterisk character (*) as a placeholder for one or more unknown peak value(s), the software will return the best match(es). The use of comparison operators, which, for instance, can display values that are greater than a desired peak value, can increase this usability. Furthermore, users can perform searchesbased on a mineral's name, che-

		Raman	spectrum of Calcite	Unk rdrs.	uaic.re/minerals/calcite.html			Open lini	
Feference: Bugger N., Appel A. I. (2001) The Flavour today on orders and testistics. Analist 60 (2016) as 1 (Striventiat 170. I. Cuar - List, Term 65, Issue 2, 17-112.									
8,500									
8,000	-				Reference inum	ner 1. Anner 5. 1. (2005) The Reman study on certain	calineates. Analais Statifics de Universitat	a "Al 1 O ca" dan Tana M.	ins at 7.
7,500	E				97-1	112.			
7,000 -	a		Capy						
0,500	-				Heb Heb	License Report		Orystal Assignments	i Pho
5,000			Save as	0					
5,500		-	Dist	-	Main Peak(s) of	Calote	Tentative Assignment		
4,500					1087		v1 symmetric stretching of CO3		
4,000			Zeom	n 21					
2,500		1							
3,000			Zeom I	Jul -					
2.500 -			Auto B			•			
2.000	1					12120	Contractor of the		
1.500			Reverse	X-8(15	Cither Pesk(s) et	rcaote	Tentative Assignment		
1,000			Hele/9	testa er	715		visionetric deformation		
500	, U V				1537		v3 asymmetric stretching		
0 -	1 mmmmulu		men have marked	Constraint and a second and a second and a second	1749		v1 + v4		
1ar search 🖲	Open Spectrum Save s	1,000 1,250 1	(,500 1,750 2,000 2,250 Wavenumber (cm ⁻¹) Nain p	2,500 2,750 3,000 3	250 3,500 - 2051; y = 7109				
ear search V	Open Spectrum Save s	1,000 1,250 1	(,500 1,750 2,000 2,250 Wavenumber (cm ⁻⁺) Nair p	2,500 2,750 3,000 3	250 3,500 (= 2051; y = 709				
ar search T nary Table Sample#	Open Spectrum Save s	1,000 1,250 1 Decturi Clear A	(500 1,750 2,000 2,150 Wavenumber (cm ⁻¹) Nain p Otenical formula	2,500 2,750 3,000 3	250 3,500 < - 2051 : y - 7109 Other peaks (on *9)	Reference	Unk	Comments	
ar search 🕈 nary Table Sample#	Open Spectrum Save s	1,000 1,250 1	(500 1,750 2,000 2,250 Wavenumber (cm*1) Rain p Ovenical Formula	2,500 2,750 3,000 3	250 2,500 - 2051 : y - 7199 Other peeks (cn ⁻¹)	Reference	Unk	Connexts	
ar search V nery Table Sample# M13	Open Spectrum Save s	1,000 3,250 3 pocture Clear A Mineral Mineral	(500 1,750 2,200 2,250 Wavenumber (cm*1) Nain p Otencel formula	2,500 2,750 2,000 3	250 3,500 c = 2051; y = 7299 Other peeks (on *) 55,715, 107, 1949	Reference Bager (n. Ascent A. J. (2000) The Ram	Unik v j das useči najmengra je jedicita žiteli	Comments	
ar search T nary Table Sample# S455 S455	Open Spectrum Save 1 Open Spectrum Save 1 Open Spectrum Cess Carbonates/Mitoles/Orales Carbonates/Mitoles/Parales Carbonates/Mitoles/Parales	1,000 2,250 2	(300 1,750 2,000 2,250 Wavenumber (om ") Reir p Orencefromia 2,550) Perf00,3 Perf00,3	2,500 2,750 3,000 3	250 3,500 (= 2051 ; y = 7199 Other peaks (on *) 505,715, 1022, 144 280,545,727, 1452, 128 291 392 (upo 192)	Reference Name NL, Associa A, L. (2005) The Time Barger NL, Associa A, L. (2005) The Time Barger NL, Associa A, L. (2007) The Time	Link o da uart of hans an specialize hand o da uart of hans an specialize hand o da uart of hans an specialize hand man hand hand hand hand hand hand hand ha	Comments	
ar search V nary Table Sample# 2433 2455 2455 2455 2455 2455	Open Spectrum Save 1 Open Spectrum Save 1 Open Spectrum Open Open Spectrum Open Ope	1,000 2,250 3	L00 1/30 2,000 2,250 Wavenumber (cm.*) Kein p Orenical Funda Coenical Funda	2,500 2,750 3,000 3	250 2,500 (= 2551; y = 709 (http://doi.org/) (5,70; 407, 128 (20,54; 77, 149, 128 (20,54; 77, 149, 128 (20,54; 77, 149, 128 (20,54; 77) 143, 145 (20,54; 77) 145 (20	Adverse Juger Nu Acces A. E. (2001) free Tata Buger Nu Acces A. E. (2001) Feb Tata Buger Nu Acces A. (2001) Feb Tata Buger Nu Acces A. (2001) Feb Tata	Link Visit uskung hinning skalata hinni Visit uskung hinning skalata hinni Visit uskung hinning skalata hinni Visit uskung hinning skalata hinning hinni	Connetts	
ar search V nary Table Sample# 5455 5455 5455 5455 5455 5455 5455 54	Open Spectrum See 3 Gen	1,000 1,200 1 Decturi Clear A Mineral Celore Sectore	.00 1,70 2,00 2,20 2,20 Wavenumber (em.*) Wavenumber (em.*) Genoaf formia Ocensal formia Costo) Re ¹ r(co) Re ² r(co) Costo)	2,500 2,750 3,000 3	250 2.500 (= 2051 y = 709 00Her peaks (px*9) 55,711, 1402, 124 285,54,727, 1402, 127 287, 298, 494, 577, 702, 1251, 1255 297, 209, 1401, 1273	Reference Nation M, Apope A I, Coliffy The Nem- Bauger M, Apope A I, Coliffy The Nam- Bauger M, Apope A I, Coliffy The Nam- Bauger M, Apope A I, Coliffy The Nam-	Lok organistic program (calcito the organistic program) (calcito t	Connecto	
ar search T mary Table Sample# 5413 5455 5433 Synthetic 5473 5404	Open Spectrum See 1 Open Spectrum See 1 Open Spectrum See 1 Open Spectrum Open Open Spectrum Open Open Spectrum Open	1,000 2,250 2 Doctum Clear A Some Rhockhoste Aragonia Website		2,500 2,750 3,000 3 with 1087 cm1 Majas dick cm (250 2.500 Clifer peoks (m *) 507,723,135,723 208,723,145,723,145 207,264,947,725,145 207,264,947,725,145 207,264,947,720,1354,145 207,951,1452,2131	Roberton August A. Appent A. L. (2001) Than Anne Bauger M., Appent A. L. (2001) Than San Bauger M., Appent A. L. (2001) The San Bauger M., Appent A. L. (2001) The San Bauger M., Appent A. L. (2002) The San	Link data saka riya hara sa kasi kati kana rosa saka sahara saka kati kana rosa saka sahara saka saka kati kana rosa saka sahara saka saka saka saka rosa saka sahara saka saka saka saka rosa saka sahara saka saka saka saka saka rosa saka saka saka saka saka saka saka s	Corrects	
ar search T mary Table Sample# 5455 5455 5455 5464 Synthetic 5404 Synthetic	Open Spectrum Sove 1 Open Spectrum Sove 1 Open Spectrum Open Ope	1.00 1.00 3 Boctur Cear 4 Meerd Sdork Bockrive Bockrive Meerd Mee	0.00 1,730 2,000 2,250 Wavenumber (em ') Rec p decoy Re ² (co) Re ² (co)	2,500 2,750 3,000 3 wiki (1937 cm 1) Waara did cm (0) Mer peats (cm 1) 1957 19	250 2,500 - 35111 - 739 Offer peels (or 1) 50,711, 402, 249 240, 541,727, 1492, 259 251,728, 4195, 131 257, 264, 465, 131 257, 264, 104, 151 257, 264, 104, 151 257, 264, 104, 151 257, 264, 104, 151 257, 264, 104, 154, 109 257, 264, 104, 154, 109 257, 264, 104, 154, 109 257, 264, 104, 154, 109 257, 254, 154, 154, 154 257, 254, 154, 154 257, 254, 154, 154 257, 254, 154, 154 257, 254, 154 257, 254, 155, 154 257, 254, 154 257, 254, 155, 154 257, 254, 155, 154 257, 254, 155, 154 257, 254, 155, 154 257, 254, 154 257, 254 257, 257 257, 25	Reference Papar R., Josep R. J. (2005) The Man. Bargar K., Josep R. J. (2005) The Man.	Lonk Colouria California de California de California Colouria de California de California de California Colouria de California de La California de California Colouria de California de California de California Colouria de California de California de California Colouria de California de California de California California de California de California de California de California de California de California de California de California de California de California de California de California de Calif	Connerts	
ar search V nary Table Sample# 5455 5455 5455 5454 5474 5494 5494 5494	Copes Spectrum Gene	1.00 1.00 2 sector Cear A More d Sector Sector vala Cregorite Witherite Sector		2.500 2,750 3,000 3 with (1877 cm) Mouse doit cm (c) More peets (cm ?) 1022 1023 1024 1025 1025 1026 1026 1027 1026 1027 10	250 2.500 Clifer peols (m *) 507 (107 2.40) 508 (117 - 7109 509 (117 - 7109 509 (117 - 710) 509 (117 - 710) 500 (117 -	Before c Barrier, Andre A. L. (2000) The family Bagratin, Angre A. L. (2000) The family Bagratin, Angre A. L. (2000) The family Barrier, Angre A. L. (2000) The family	Likk Costa da Dina marka da Bana costa da Dina marka da Bana costa da Dina marka da Bana costa da Dina da Dina da Dina costa da Dina da D	Correcto	
ar search V Sample# 5455 5453 5455 5453 5464 5473 5464 5473 5464 5473 5464 5473 5464 5473 5464 5473	Constructions Construction	1.00 1.00 2 contrue Cear A Serie Backing Serie Backing Cears Cea	2.00 1,750 2.00 2,250 Wavenumber (m ⁻¹) Wavenumber (m ⁻¹) Genosif funda Genosif funda Genosif funda Genosif funda Genosif funda Genosif Fieldo Genosif Fieldo Fie	2,250 2,270 2,000 2 ali (1970m) Hean dolon (0) Hean petro (m-1) 107 107 107 107 107 107 107 107	200 3.500 (1-2011) y - 700 (0) pole (0) y - 700 (0) 51, 140, 276 (0) 51, 140, 276 (0) 51, 140, 276 (0) 51, 140, 276 (0) 51, 140, 176 (0) 50, 141, 120 (0) 51, 141, 170 (0) 51, 141, 141 (0) 51, 141, 141 (1) 51, 141, 141, 141, 141 (1) 51, 141, 141, 141, 141, 141, 141, 141,	Seferrer: Seferrer: Seferrer: Regarin, Angel A. L. 2005 The tam. Buggrin, Angel A. L. 2005 The tam.	Units Consume Log American California Consumer Log American California Consumer Log American California Consumer California Consum	Connerts	
ar search V Sample # Sample # Set13 Set33 Set34 Synthetic Set33 Set04 Synthetic Set33 Set04 Synthetic	Genes Spectrum Genes Spectrum Genes Genes Spectrum Genes Gene	1.00 1.00 2 sectur Cear A Sector Se	500 1750 200 2.05 Wavenumber (m1) More Perfox Perf	2,200 2,270 3,000 3 with 1987 on 1 House dot em (6 1) Mar peeks (on 1) 100 100 100 100 100 100 100 1	200 2,500 00% 20% (0,1%) 20%	Sofeware Margarit, Assard A.B. 2003 Peril Regarit, Regarit, B.B. 2003 Peril Regarit, Regarit, B.B. 2003 Peril Regarit, Regarit, B.B. 2003 Peril Regarit, Regarit, B.B. 2003 Peril Regarit, Regarit, R.B. 2003 Peril Regarit, Regarit, B.B. 2003 Peril Regarit, Regarit, R.B. 2003 Peril Regarit, Regarit, Regarit, R.B. 2003 Peril Regarit, Regarit, Regarit, Regarit, R.B. 2004 Peril Regarit, Regarit, Regarit, Regarit, Regarit, R.B. 2003 Peril Regarit, Regarit,	Lok estatutional and a second and a second estatutional and a second and a second and a second estatutional and a second and a second and a second estatutional and a second and a second and a second and a second estatutional and a second and a second and a second and a second estatutional and a second and a second and a second and a second and a second estatutional and a second and a	Connects	
ar search V mary Table Sample# 5455 5458 5458 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5475 5475 5475 5475 5475 5475 5475	Cyens Spectrum Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Censul 1 Centron 1 Ce	1.00 1.00 2 contrain Cear a Neeral Sofert Boack/vala Cegrynia Without Soferts Sofert		2,200 2,270 3,000 3 2,200 2,270 4,000 4 doi: 10 1000 000 000 1 1000 000 000 0 1000 000 0 1000 0	200 2,500 (00er goals (srv 7) 00er goals (sr	Advance Jougn K., Augus A. J. (2015) The Kan- Bargh Ta,	Like - Standard Johnson Like Johnson - Standard Johnson -	Connerts	
ar search 1 mary Table Sample# Set55 5455 5455 5454 5454 5454 5454 5454	Genes Spectrum Genes Spectrum Genes Spectrum Genes Genes Spectrum Genes	1.00 1.00 3 Destuin Crear A Reard		2,200 2,700 3,000 3 with 100 mm ⁻¹ (Mone statum (g. * *)) 1 200 200 2 200 200 200 200	200 3.50 (************************************	Selectors Compared and an analysis of the selectors Regards, Appendix A. B. (2010) The selectors Regards, Appendix A. (2010) The selectors Regards, Ap	Lok. Casada Jones (Jones Jones) Casada Jones (Jones) Casada Jones Casada Jones (Jones) Casada Jones Casada Jones Casa	Correcto	
ar search Sample# Sample# 24555 2455 2455 24555 24555 24555 24555 24555 245555	Constraints of the sector	1.00 1.00 2 section Cerr A Cerr A C	200 129 200 200 200 200 200 200 200 200 200 2	2,250 2,270 3,000 3 with 1597 on 1, Means did one (2 Main peaks (per 7) Main peaks (peaks (peaks (peaks (peaks (peaks (peak	202 2,500 608 (pr. 1) 608 (pr	Reference Bauger L, Aport AL, DOND The Mail Bauger L, Aport AL, Aport AL, DOND The Mail Bauger L, Aport AL, DOND The Mail Bauger L, Aport AL, DOND The Mail Bauger L, Aport AL, DOND The Mail Bauger	Lok Constantial Constantia Constantial Constantia Constantia	Correcto	
ser search V Sengle# Sets Synthetic Sets Sets Sets Sets Sets Sets Sets Sets	Constraints of the second	1.00 1.00 1 Bentin Dear A Status St		2,250 2,70 3,000 3 with (M27 mm1) Means side on [0] 1 acc acc 1 1 acc 1 1	200 2,500 - 2001 ; 1 - 7,00	Before Participant Partipant Participant	DA Common characterization from Common characterization for Common characterization Common c	Connects	
nr search V Sample# A12 A155 A135 A155 A133 A155 A133 A157 A133 A157 A133 A157 A157 A157 A157 A157 A157 A157 A157	Constraints of the second	1.00 LOS 1 Corr A Sector	200 1.20 2.00 Wavemake (c.f.) Kep Reconstruction Reconstruction Reconstreconstruction R	2,250 2,700 3,000 3 with 10,870 m²l (Mood dat m²l (0 ···²)) 1 1 Mare partie (m··²) 1 1 1 Mare partie (m··²) 1 1 1 1 Mare partie (m··²) 1 <td>200 3,500 - 2021 y - 7,00 Deep code (pr 7) 50,144,103,100,144 20,244,105,100 20,244,100</td> <td>Schemer: Security and an annual schemer and annual schemer Regard M, Anger A E. (2003) The flat. Regard M, Anger A E. (2003) The fla</td> <td>DA Contract Contract Contract Contract Contract C</td> <td>Connects</td> <td></td>	200 3,500 - 2021 y - 7,00 Deep code (pr 7) 50,144,103,100,144 20,244,105,100 20,244,100	Schemer: Security and an annual schemer and annual schemer Regard M, Anger A E. (2003) The flat. Regard M, Anger A E. (2003) The fla	DA Contract Contract Contract Contract Contract C	Connects	
ar search V Sample# Sample# Set35 Se	La construction de la construcción de la construcci	1.00 1.00 1 Sector Control Co	2001 129 2001 239 Warenaber (r. 17) Warenaber (r.	2.200 2.700 3.000 3 with 10.97 orm Network days in (i.e., i.e., i.	20 3.50 - 3011 y - 702 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Reference Registrin, Access A. J. 2008 (De Maria) R	Lok Common	Connects	
sar search V mary Table Sample# 5455 5433 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5404 5473 5405 5473 5405 5473 5405 5473 5404 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5404 5473 5405 5473 5404 5473 5405 5473 5405 5473 5404 5473 5404 5473 5405 5473 5404 5473 5404 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 5405 5473 54755 5475 5475 5475 5475 5475	acc (Appent Speechurg) (Appent Speechurg) [Copent Speechurg) (Appent Speechurg) (Appent Speechurg) [Copen	1.00 1.00 1 Bechnik Corr A Bechnik Corr A Bechnik Corr A Bechnik Corr A Bechnik Corr A Corr A Bechnik Corr A Bechnik C		2,200 2,700 3,000 3 with 10,700 mm Nations Status () 2 700 2000 2 2 700 2000 2 2 700 2000 2 2 700 2000 2 2 700 2000 2 2 700 2000 2 2 700 2000 2 2 700 2 2 2 700 2 2 2 2 700 2 2 2 2 700 2 2 2 2 700 2 2 2 2 700 2 2 2 2 2 700 2 2 2 2 2 2 700 2 2 2 2 2 2 2 2 2 2 2 2	200 2.500 (1) (1) (1) (2) (2) (1) (1) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2	Selence Selence (1990) A 1990	10- 20- 20- 20- 20- 20- 20- 20- 2	Corrects	
ear search imary Table Sample # 5453 5453 5454 5473 5464 5473 5484 5473 5484 5485 5485 5485 5485 5485 5487 5487	La construir de la construire de la construir de la construir de la construir de la construir	1.00 LSS 1	200 1.20 2.00 2.20 2.20 2.20 2.20 2.20 2	2.200 2.270 3.000 3 with 10.2707 Neuro diam (6 7)	200 3,500 COL 19 - 7000 Other peaks (pr. 7) 500, 541, 70, 730, 500 State (pr. 10, 10, 10, 10, 10, 10, 10, 10, 10, 10,	Software Bargeris, Associa A. J. 2000 Perform Bargeris, Bartina, A. Jorder F. C. D. Bargeris, Bartina, J. Perform F. C. D.	UK Anter and Anter and Anter		

Fig. 1 Layout of the main window of the Raman Data Search and Storage (RDSS) application with a data entry selected. a) a chart panel where the Raman spectrum is interactively displayed; b) two tables tabbed with the data entries; c) specific vibrations with the tentative assignment related to each selected entry; d) properties of the Raman spectrum.

AUI-G, 65-66, (2019-2020) 5-18



Fig. 2 An example showcases the various features available in the software for visualizing a) a photo of the selected mineral entry and b) the crystal structure.

mical formula, or even mineral class. The *RDSS* program also includes tools for tasks aimed at basic data representation and analysis of the raw Raman spectrum by using the open spectrum function.

Apart from the possibility to search by the wavenumber positions of the Raman bands, this program provides users with the ability to see the specific photo of the entry reference (Fig. 2–a), to save the Raman spectrum as a PDF file or as an image (PNG), or even to copy it to the clipboard. Moreover, *Jmol* (Hanson et al., 2013) is implemented to show the interactive 3D crystal structure of the specific mineral entry (Fig. 2–b).

Figure 3 provides a summary of the relationship between the database and the various functionalities available in the software. Displaying and examining the Raman spectra are the core concepts of the

software, where peak finder and intensity assignment tools are provided to analyze the spectra.

3. Program description

3.1. Data entries

The *RDSS* program has 84 Raman entries, which include minerals from carbonates (Buzgar and Apopei, 2009), sulfates (Buzgar et al., 2009), pyroxenes (Buzatu and Buzgar, 2010), amphiboles (Apopei and Buzgar, 2010), and other mineral classes or groups.

Data entries called from the web server are stored in two tables. Each of them is part of a tabbed panel component. The first tab, called the "Primary Table," contains a table designed to have the following common columns: ID, sample number,



Fig. 3 Simplified workflow of the Raman Data Search and Storage (RDSS) program.

mineral class, mineral name, chemical formula, main peaks, other peaks, reference, external link, and comments.

The second tab, called the "Secondary Table," contains a table designed to have more specific columns for a logical search for peaks. The columns are structured as follows: mineral name, 200-600 cm⁻¹, 600-900 cm⁻¹, 900-1200 cm⁻¹, 1200-1500 cm⁻¹, 1500-2000 cm⁻¹, 2000-3000 cm⁻¹, and 3000-4000 cm⁻¹ spectral domain, external link, comments, and reference. The first and last three columns are linked to those of the "Primary Table" for logical The fundamental vibrational reasons. modes of the covalent chemical bonds in the crystal structures (such as SO₄, PO₄, CO₃, etc.) determine the Raman spectral patterns and the location of the main Raman bands, which is the reason for the splitting into multiple spectral domains.

Apart from these structured tables, other tables are stored within the database,

namely XY datasets for the Raman spectra, vibrations and tentative assignments of each mineral species, photos, and crystal structures.

and they contain other types of data,

3.2 Supported file formats

The RDSS application can open any file with the following extensions: txt, rruff, csv, asc, dat, or any XY data format, where the data in the first column is assumed to be the wavenumber and those in the second column designate intensity. The RDSS program supports the following file format delimiters: (i) tab delimiter; (ii) space; (iii) comma; (iv) semicolon; and it can even ignore rows that start with "##" (as a comments). These delimiters increase the possibility of opening spectra that are formatted in different ways (this occurs when different types of spectroscopic instruments and/or standardization are used). For example, the files of the RRUFF project contain a two-digit sign (##) for each line that includes the name, locality, ID, and chemistry of the sample. Proprietary file formats are not supported, but many of them have the option to export in the above-mentioned plain format.

3.3. Plotting and visually exploring the Raman spectra

The RDSS application supports two methods for opening raw Raman spectra: (i) through the standard "File" menu or (ii) by pressing the "Open Spectrum" button from the main window. Both of them use the File Chooser dialog to browse for the desired files. The Raman spectra called from the server database and those that are opened from the local machine can be visually explored in different ways, thus providing a series of functionalities. They include the zooming and rescaling of the Raman spectrum and the dragging, peaking, printing, and showing of the coordinates as the mouse moves over the Raman spectrum. A list of the actions that can be performed on the Raman spectrum is provided in Table 1.

The chart panel where the Raman spectrum is shown has functionalities and properties that can be activated through the actions of the second-click mouse. The right click activates a popup menu (Fig. 1- d) with the items listed in Table 2.

3.4. Raman Data Search and Storage (RDSS) tools

3.4.1. Analyzing procedure

This allows users to manually select the location of the peaks stored in a peak list that is used for comparison with the database peak values from the main window. Furthermore, an automatic algorithm for the finding of peaks is implemented. This is a great time-saving tool when it comes to Raman spectra with several peaks (Fig. 4).

When the "Manual" radio button is pressed, a panel with the manual search function becomes active for the location of Raman peaks. After the "Get Peaks" checkbox is checked, on the Raman spectra chart panel, a blue crosshair appears, and it moves along with the mouse crosshair. In other words, it is similar to real-time mouse tracing, which helps the user obtain the exact peak position. The user can also zoom in before obtaining the desired peak.

When the "Automatic" radio button is pressed (Fig. 4–a), a panel with the auto-

Action	How to perform it
Zoom in/out	use the scroll wheel up/down
Zoom in on a desired region	use the left mouse button to drag a selection rectangle
Restore X-axis / Y-axis	press the right mouse button and drag the mouse up/down
Translate X-axis / Y-axis	Ctrl + left mouse button and drag the mouse to move the whole Raman spectrum
Get X value	hover the mouse on the Raman spectrum to see the specific value of the X axis

Tab. 1 Mouse actions that can be performed on the Raman spectrum

Item	Action
Сору	copy the Raman spectrum (you can paste it in photo editors)
Save as	save the Raman spectrum in the PNG file format
Print	print it
Zoom in > Both Axes	zoom in preserving the ratio of both axes
> Domain Axis	zoom in preserving only the intensity ratio
> Range Axis	zoom in preserving only the wavenumber ratio
Zoom out > Both Axes	zoom out preserving the ratio of both axes
> Domain Axis	zoom out preserving only the Intensity ratio
> Range Axis	zoom out preserving only the wavenumber ratio
Auto Range > Both Axes	auto Range both axes
> Domain Axis	auto Range wavenumber axis
> Range Axis	auto Range intensity axis
Reverse X-axis	reverse wavenumber axis
Help/Shortcuts	opens a window with instructions on how to use the mouse for the fast and easy control of the Raman spectrum

Tab. 2 Functionalities and properties of the Raman spectrum



Fig. 4 Automatic peak finder operation used on a high-fluorescence Raman spectrum by setting the threshold (b) and gap variables.

matic search function becomes active for the location of Raman peaks. In this case, the algorithm is quite simple: a tangent (a line between two adjacent points) is

extended to every point of the dataset, after which the program looks for changes in the sign of the slope of the lines (a simplified version of finding zeros in the first derivative of a function). To use the automatic search function, the user needs to fill in two variables before obtaining peak values. The first variable is the "threshold" (Fig. 4-b). By specifying the threshold height, "findpeaks" only returns peaks that exceed at least the Y-value of the "threshold". The second variable is the "gap." By specifying the gap width, this function affects a number of peaks, as the density of the peaks is directly correlated with this variable. The recommended values are between 50 and 300, but, in some cases (due to the high density of sharp peaks), it is better to play around with these two variables.

Regardless of the method used (manual or automatic) to extract the location of the peaks, after at least one peak is found, all of them appear in a list with the following options: delete a single peak, multiple peaks, or even all peaks from the list. To compare the selected peak(s) from the list with the database, press the "Go" button from the bottom of the entire panel.

3.4.2. Relative Intensity Assignment (RIA)

This action is performed on both baseline-adjusted Raman spectra and Raman spectra with high fluorescence. In both cases, five levels of intensity assignment are used, as shown in Table 3.

Baseline-adjusted Raman spectra are designed to work with baseline-removed spectra. After the "On/Off | Intensity Assignment" checkbox is checked, the user must check the "Strongest peak" radio button, and then extract the strongest peak by moving the crosshair and pressing the first mouse button on that particular peak. Afterwards, the lowest peaks (in terms of intensity) can be assigned based on the strongest peak, following the same steps (Fig. 5).

High fluorescence Raman spectra, on the other hand, are designed to work with the Raman spectra that are not baselineremoved (the spectrum shows a characteristic curvature). After the "On/Off | Intensity Assignment" checkbox is checked, the user must extract the strongest peak in two steps, as follows: (i) First, they must identify the highest peak from the Raman spectrum and check the

Intensity	Assignment
between 0 – 10%	very weak (vw)
between 10 - 30%	weak (w)
between 30 – 60%	medium (m)
between 60 – 90%	strong (s)
between 90 - 100%	very strong (vs)

Tab. 3 Steps in the intensity-based assignment of Raman bands



Fig. 5 Raman intensity assignment panel for a baseline-adjusted spectrum and the steps taken in the assignment of all peaks in relation to the strongest Raman band.

"Min" (which stands for "minimum") radio button, and then extract the minimum of the strongest peak by moving the crosshair and pressing the first mouse button on that minimum. (ii) The second step is similar to the first; the difference lies in the fact that the user must check the "Max" (which stands for "maximum") radio button and then extract the maximum of the strongest peak by moving the crosshair and pressing the first mouse button on the maximum of the peak. For the assignment of Peak 1, the procedure is similar to that of the extraction of the intensity value for the "strongest peak" (discussed above).

4. Testing

In order to test the identification feature, a raw Raman spectrum was acquired. A comparative test using the same raw dataset was carried out using the *CrystalSleuth* program (Laetsch and Downs, 2006), which is a tool derived from the RRUFF project (Downs, 2006). The test involves a Raman spectrum with few spectral bands and high fluorescence (the spectrum shows a characteristic curvature).

Fluorescence is, generally, due to the presence of organic material, but it can also be observed in inorganic materials, in which case it can be attributed to the inadequate laser wavelength used, the presence of impurities, the ambient temperature, or the temperature at the surface of the sample (if a high-power laser beam is used). Fluorescence often causes difficulties when using similarity search algorithms, which treat the spectrum as a vector. Utilizing baseline removal and smoothing as part of pre-processing techniques can solve the issue. These techniques are,



Fig. 6 Running the search algorithm by using the value of the Raman band (a) extracted from the raw Raman spectrum as a comparator (b) with the database. The results are filtered, and the corresponding Raman spectrum (c) is shown after the desired result row is clicked.

however, time-consuming and not recommended for non-experienced users, while the use of this program is fast and user-friendly. The user must first search for the peaks in the raw Raman spectrum by using the analyze tab panel. Two options are available for the identification of the spectral bands: manual and automatic, and they have been discussed above (Fig. 4). Instead of using a preprocessing method, which, as previously stated, consists of baseline removal and, in some cases, smoothing, on the raw Raman spectrum, one can use a simple similarity search by peak positions. The true value of this piece of software is shown when the user wants to search by the main peak or by other peaks using the green fields from the specific column of the primary table (Fig. 6). They can also

search for peaks between seven spectral domains using the same green fields in the secondary table tab.

Using the same raw dataset that was found by the CrystalSleuth program (Laetsch and Downs, 2006) for a comparison test (Fig. 7) shows that the bestmatching results do not agree well with the input dataset. Moreover, the search operation is quite slow (around 30 seconds for this dataset). It is worth noting that the entire GUI freezes until the search is complete. For the RDSS application, the search finishes in less than 1 second (almost instantly). One must, however, take into account that the two programs use different strategies for the identification process. For both test scenarios, an Inte(R) Core(TM) CPU (model i7-2630QM at 2 GHz) with 8 GB of memory (RAM) was used.



Fig. 7 The usage of the *CrystalSleuth* program (Laetsch and Downs, 2006) for a comparative test purpose. In comparison to the input dataset (the Raman spectrum of Quartz), the best-matching results show poor agreement (i.e., Amber).

Future work

Future work will concentrate on populating the reference database with new mineral classes. In addition, some methods for the pre-processing of the raw Raman spectrum need to be implemented (e.g., normalization, smoothing, and baseline removal). In terms of data entries, additional information, such as sample location, occurrence and association of each mineral entry, and optical and physical properties, is necessary. Moreover, by using the same codebase with a limited range of *RDSS* functionalities. this application may be extended as а JavaScript web application or available for mobile, allowing users to access and analyze the reference database online. Additionally, future work may also

involve incorporating machine learning algorithms to enhance the accuracy and efficiency of mineral identification based on Raman spectra.

5. Conclusion

Raman Data Search and Storage (RDSS) is a powerful Java-based program that allows for the display and analysis of Raman spectra with high efficiency and accuracy. *RDSS* program serves as a fundamental tool for researchers in Raman spectroscopy, offering a user-friendly interface and robust search features. It simplifies the process of identifying unknown minerals and improves the investigation of Raman data. Its comprehensive functionalities, including peak finding algorithms and intensity assignments, offer

a holistic solution for researchers grappling with complex spectra.

The main key features of the RDSS include:

- visualization of Raman spectra with an intuitive interface;
- advanced search functionalities based on peak positions and spectral characteristics;
- manual and automatic peak finding algorithms for spectrum analysis;
- intensity assignment tools for baselineadjusted and high-fluorescence spectra.

Availability and requirements

Project name: Raman Data Search and Storage (RDSS)

Project home page:

http://www.rdrs.ro/rdss/

Operating system(s): Platform independent

Programming language: Java

References

- Apopei, A.I., Buzgar, N., 2010. The Raman study of amphiboles. Analele Stiintifice ale Universitatii "Al. I. Cuza" – Iasi, Geologie, 56, 1, 57–84.
- Buzatu, A., Buzgar, N., 2010. The Raman study of single-chain silicates, Analele Stiintifice ale Universitatii "Al. I. Cuza" – Iasi, Geologie, 56, 1, 107–125.
- Bakker Schut, T.C., Wolthuis, R., Caspers, P.J., Puppels, G.J., 2002. Real-time tissue characterization on the basis of in vivo Raman spectra. Journal of Raman Spectroscopy, 33, 580–585.
- Bersani, D., Lottici, P.P., 2010. Applications of Raman spectroscopy to gemology. Analytical and Bioanalytical Chemistry, **397**, 2631–2646.
- Burrafato, G., Calabrese, M., Cosentino, A., Gueli, A.M., Troja, S.O., Zuccarello, A., 2004. ColoRaman project: Raman and fluorescence spectroscopy of oil, tempera and fresco paint pigments. Journal of Raman Spectroscopy, **35**, 7.

- Buzgar, N., Apopei, A.I., 2009. The Raman study on certain carbonates. Analele Stiintifice ale Universitatii "Al. I. Cuza" – Iasi, 55, 2, 97–112.
- Buzgar, N., Apopei, A.I., Buzatu, A., 2009. Romanian Database of Raman Spectroscopy (http://rdrs.uaic.ro). Accessed on September 2013.
- Buzgar, N., Buzatu, A., Sanislav, I.V., 2009. The Raman study on certain sulfates. Analele Stiintifice ale Universitatii "Al. I. Cuza" – Iasi, 55, 1, 5–23.
- Castanys, M., Soneira, M.J., Perez-Pueyo, R., 2006. Automatic Identification of Artistic Pigments by Raman Spectroscopy Using Fuzzy Logic and Principal Component Analysis. Laser Chemistry, 2006, 1–8.
- Downs, R.T., 2006. The RRUFF Project: an integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals. Program and Abstracts of the 19th General Meeting of the International Mineralogical Association in Kobe, Japan. O03–13.
- Hanlon, E.B., Manoharan, R., Koo, T.W., Shafer, K.E., Motz, J.T., Fitzmaurice, M., Kramer, J.R., Itzkan, I., Dasari, R.R., Feld, M.S., 2000. Prospects for in vivo Raman spectroscopy. Physics in Medicine and Biology, 45, R1–59.
- Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., Sussman, J.L. 2013. JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. Israel Journal of Chemistry, 53, 3-4, 207–216.
- Ishikawa, S.T., Gulick, V.C., 2013. An automated mineral classifier using Raman spectra. Computers & Geosciences, 54, 259–268.
- Laetsch, T., Downs, R.T., 2006. Software for Identification and Refinement of Cell Parameters from Powder Diffraction Data of Minerals Using the RRUFF Project and American Mineralogist Crystal Structure Databases. Abstracts from the 19th General Meeting of the International Mineralogical Association, Kobe, Japan, 23–28 July 2006.
- Lindholm, T., Yellin, F., Bracha, G., Buckley, A., 2012. The JavaTM Virtual Machine Specification. Java SE 7 Edition.
- Lowry, S., Wieboldt, D., Dalrymple, D., Jasinevicius, R., Down, R.T., 2009. The Use of a Raman Spectral Database of Minerals for the Rapid Verification of Semiprecious Gemstones. Spectroscopy, 24, 7.

- Nasdala, L., Smith, D.C., Kaindl, R., Ziemann, M.A., 2004. Raman spectroscopy: analytical perspectives in mineralogical research. EMU Notes in Mineralogy, EMU School on Spectroscopic Methods in Mineralogy, 7, 281–343.
- Perez-Pueyo, R., Soneira, M.J., Ruiz-Moreno, S., 2004. A fuzzy logic system for band detection in Raman spectroscopy. Journal of Raman Spectroscopy, 35, 808–812.
- Popp, J., Krafft, C., Mayerhöfer, T., 2011. Modern Raman spectroscopy for biomedical applications. Optik & Photonik, 6, 4, 24–28.
- Reisner, L.A., Cao, A., Pandya, A.K., 2011. An integrated software system for processing, analyzing, and classifying Raman spectra. Chemometrics and Intelligent Laboratory Systems, **105**, 83–90.
- Salimbeni, R., Castanys, M., Soneira, M.J., Perez-Pueyo, R., Ruiz-Moreno, S., Pezzati, L., 2005. A fuzzy logic system for Raman spectrum identification. Proceedings SPIE 5857, Optical Methods for Arts and Archaeology, 585714 (12 August 2005); https://doi.org/10.1117/12.612608.
- Sharma, S.K., Lucey, P.G., Ghosh, M., Hubble, H.W., Horton, K.A., 2003. Stand-off Raman spectroscopic detection of minerals on planetary surfaces. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 59, 2391–2407.
- Soneira, M.J., Perez-Pueyo, R., Ruiz-Moreno, S., 2002. Raman spectra enhancement with a fuzzy logic approach. Journal of Raman Spectroscopy, 33, 599–603.

- Vandenabeele, P., 2004. Raman spectroscopy in art and archaeology. Journal of Raman Spectroscopy, 35, 607–609.
- Vandenabeele, P., 2011. Evaluation of a spectral searching algorithm for the comparison of Raman band positions. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, **80**, 27–31.
- Vandenabeele, P., Tate, J., Moens, L., 2007. Nondestructive analysis of museum objects by fibreoptic Raman spectroscopy. Analytical and Bioanalytical Chemistry, **387**, 813–819.
- Vankeirsbilck, T., Vercauteren, A., Baeyens, W., Van der Weken, G., Verpoort, F., Vergote, G., Remon, J.P., 2002. Applications of Raman spectroscopy in pharmaceutical analysis. TrAC Trends in Analytical Chemistry, **21**, 869–877.
- Virkler, K., Lednev, I.K., 2010. Forensic body fluid identification: the Raman spectroscopic signature of saliva. Analyst, **135**, 512–517.
- Wang, A., 2003. Development of the Mars microbeam Raman spectrometer (MMRS). Journal of Geophysical Research, Planets, **108**, E1, 1–18.
- White, S.N., Dunk, R.M., Peltzer, E.T., Freeman, J.J., Brewer, P.G., 2006. In situ Raman analyses of deep-sea hydrothermal and cold seep systems (Gorda Ridge and Hydrate Ridge). Geochemistry, Geophysics, Geosystems, **7**, 5, 1–2.

Published: March, 2024