5_JCB_Turnitin

by Jolene Hardwick

Submission date: 20-Jun-2025 08:59AM (UTC-0400)

Submission ID: 2685084984

File name: 5_Production_395__Budiman.pdf (981.98K)

Word count: 3351

Character count: 19303

The Use of C4.5 Algorithm for Classification of New Student Admissions in Vocational High Schools

(Case Study: Piramida Rancaekek Vocational High School)

Dheni Apriantsani Budiman¹, Sri Rahayu Indiani² STMIK Mardira Indonesia, Bandung ^{1,2} dheni@stmik-mi.ac.id¹, srirahayu200207@gmail.com²

Abstract

The New Student Admission is a mandatory process conducted by all educational institutions to maintain the continuity of student cohorts within the educational framework. At Vocational High School Piramida, The New Student Admission process adheres to a traditional methodology, resulting in subjective and protracted selection procedures. Consequently, the author employs the C4.5 algorithm as a classification technique that produces decision trees from potential student data. The created system is intended to manage The New Student Admission process with greater efficiency and objectivity. This study utilizes descriptive analysis, and the system development adheres to Object-Oriented Software Engineering (OOSE) principles. The system architecture emp 10 the PHP programming language in conjunction with the Laravel framework and a MySQL database. This study demonstrates that utilizing the C4.5 algorithm can enhance the efficiency and effectiveness of the student selection process, reduce errors, expedite data administration, and lead to a more objective student selection process. The method enhances information accessibility for prospective students, consequently augmenting its attractiveness to candidates.

Keywords: System, Classification, C4.5 Algorithm, Selection

INTRODUCTION

The New Student Admission is a procedure for enrolling students in educational institutions, encompassing both formal and non-formal settings. The New Student Admission is a mandatory procedure that every educational institution must follow without exception. The New Student Admission process functions as an educational ecosystem that preserves the continuity of student cohorts, which progressively evolves.

The guidelines for the execution of The New Student Admission are explicitly defined in the Ministry of Education and Culture Regulation No. 51 of 2018. In the subsequent year, these restrictions were amended in the Ministry of Education and Culture Regulation No. 44 of 2019. This refinement underscores the necessity of this activity for all educational institutions, particularly schools.

Vocational High School Piramida Rancaekek is one of the institutions that implement The New Student Admission procedure following the conclusion of annual examinations each year. The New Student Admission procedure at Vocational High School Piramida Rancaekek follows a conventional method, where prospective students, accompanied by their parents, attend the school to finalize their initial registration, submit the requisite supporting documents, and participate in essential assessments. The outcomes of new student admissions must be more objective and efficient to conserve time, as the existing approaches can yield subjective results and protracted processes.

The New Student Admission system at Vocational High School Piramida Rancaekek utilizes a classification method through the application of the C4.5 algorithm. The C4.5 method generates decision trees from categorization samples. This can be summarised as a decision tree methodology that converts extensive datasets into a decision tree that encapsulates rules.

Prior investigations into data mining in PPDB were conducted by Amelia Yusnita, Siti Lailiyah, and Khoiri Saumahudi, who employed the Naive Bayes method. The researchers determined that the decision support system employing this strategy can facilitate the student selection process in an entirely objective, impartial, and transparent manner in decisionmaking. A study was undertaken by Linda Monizah Fitriani and Andik Setyono, entitled "Application of the C4.5 Algorithm for New Student Admissions at SD Islam Terpadu Permata Bunda Demak." This study employed RapidMiner software as a modeling tool to build rules for classifying incoming student admissions. The application, designed in the Java programming language, effectively assisted the school in identifying new pupils based on their specific requirements .

Data Mining

Data mining is the procedure of gathering and analyzing data to derive important insights. This is accomplished by software that uses computations, statistics, or artificial intelligence (AI). Multiple factors are driving the eggoing evolution of data mining, including the rapid proliferation of data sets, the storage of data in warehouses that enable nearly all companies to access high-quality databases, and the competitive pressures within the business sector

to enhance market dominance in the global

C4.5 Algorithm

The C4.5 algorithm is an enhancement of the ID3 algorithm, which is utilized for constructing decision trees. CART, C4.5, and ID3 are algorithms employed in the construction of decision trees. (Chen, 2021; Tempola et al., 2022) In the work by Anwar (2025), titled "Application of K-Means and C4.5 Algorithms for dropout risk prediction in vocational high schools," Yan (2022) elucidate the fundamental methodology for constructing a decision tree utilizing the C4.5 algorithm. This approach initiates by designating an attribute as the root, thereafter forming branches for each value. Cases are subsequently allocated to these branches, and the process is reiterated for each branch until all cases within a branch are classified into the same category.

The creation of a decision tree using the C4.5 method Suherman et al., (2023) involves multiple stages:

- Compiling training data. Training data often originates from previous data that has been categorized into specific classifications.
- 2. Determining the root of the tree: The supplementary value of each characteristic is computed, with the highest supplementary value designated as the primary root. The supplementary entropy value is subsequently computed using this formula to determine the entropy value:

$$Entropy(S) = \sum_{i=1}^{n} -p_{i} * log_{2}(p_{i})$$

Description:

S = Set of cases

n = Number of partitions S

pi = Proportion of Si to S

Determine the gain value with the subsequent equation:

$$Gain\left(S,A\right) = Entropy\left(S\right) - \sum_{i=1}^{n} \frac{|S_{i}|}{|S|} * Entropy\left(S_{i}\right)$$

Description:

S = Case set A = Attribute

n = Number of attribute partitions A

|Si| = Proportion of Si to S

|S| =Number of cases in S

- Reiterate steps 2 and 3 until all records are partitioned.
- The decision tree splitting operation will cease when:
 - All records in node N belong to the same class.
 - The partitioned records no longer include any properties.
 - c. The vacant branches contain no records.

Classification

The term "classification" derives from the word "classis," which means "to group similar objects" and "to separate dissimilar objects." The classification process is based on four components (Sari & Safii, 2021): Class: This is the dependent variable that is categorical and indicates the "label" of the object. Predictor: This is the independent variable represented by the characteristics (attributes) of the data. Training Dataset: This is a set of data containing values from the two components mentioned above, used to determine the appropriate class based on the

predictor. Testing Dataset: This contains new data that the developed model will classify, and the accuracy of the classification is evaluated. Based on the explanations provided, classification is an activity performed to group or select items that share similarities in order to form a common group.

Framework

According to Yudhanto & Prasetyo, a framework is a collection of libraries or functions that developers may utilize to make it easier for them to code. (Lintang et al., 2022)

Laravel

Laravel is an open-source PHP framework distributed under the MIT license and constructed on the Model-View-Controller (MVC) paradigm. (Abrori et al., 2022) It is a PHP-based web development framework that utilizes the MVC architecture, designed to enhance software quality and minimize maintenance costs. Taylor Otwell created Laravel, recognized as one of the premier PHP frameworks. It is an open-source web development framework characterized by its attractive syntax and expressive code. Laravel is engineered to expedite and streamline web development. (Daulay et al., 2022)

15 UML

The Unified Modelling Language (UML) is a dependable modeling tool utilized in the creation of object-oriented systems. (Sulistiani & Aldino, 2020) UML is an amalgamation of modeling languages created by Booch, Object Modelling Technique (OMT), and Object-Oriented Software Engineering (OOSE). This approach renders the analysis and design process iterative,

encompassing the identification of classes and objects, the elucidation of the semantics of the relationships among these entities, and the specification of interfaces and implementations (Aji & Sunyoto, 2020; Muhammad & Rahardja, 2021).

The author aims to investigate the implementation of the C4.5 algorithm for classifying new student admissions at Piramida Rancaekek Vocational School.

METHOD

Research methods

The author utilized two primary approaches for data collecting in this study: observation and interviews. Observation involves collecting data through scrutiny of subjects in their natural habitat, allowing the researcher to document genuine behaviors and interactions. Conversely, interviews occur when the researcher gathers data by posing questions to various individuals, directing inquiries toward a set of participants to elicit their responses. Collectively, these methodologies offer an exhaustive framework for comprehending the research subject.

System Development Methods

Object-Oriented Software Engineering (OOSE) is an object modeling methodology established by Ivar Jacobson in 1992. It is the inaugural object-oriented design technique to employ use cases to enhance software design. A use case delineates a scenario illustrating user interactions with the product or system under development. This approach emphasizes use cases. OOSE has three fundamental stages: the development of requirement and analysis

models, design and implementation, and model testing. The value of this method lies in its ease of learning, as it encompasses all elements of software engineering and utilizes straightforward notation.

RESULTS AND DISCUSSION

C4.5 Algorithm Analysis

It serves for categorization or categorization and possesses predictive characteristics. The C4.5 algorithm is a data mining technique that classifies data, designed to uncover significant patterns within large datasets. The execution of the C4.5 algorithm for ascertaining new student admittance adheres to the following procedures:

1. Literature Study

At this juncture, the concepts requisite for the research will be elucidated, including an exposition of the C4.5 algorithm and its implementation methodology.

2. Data Sample Collection

At this juncture, the procedure for gathering sample data will involve soliciting information regarding potential new students for the 2024/2025 academic year from Vocational High School Piramida Rancaekek. This prospective student dataset encompasses diverse information, including names, originating schools, parental names, selected majors, and more attributes to be analyzed using the C4.5 algorithm. The overall dataset comprises 133 records of prospective pupils. Table 1 below displays several sample data entries that the author will present in this publication.

Table 1. Data on Prospective Students for Vocational High School Piramida 2024/2025

Prospective Students	Interest-Aptitude Test	Psychometric Test	Numerical Test	Major Interests	Status
Student 1	Competent	Quite Competent	Quite Competent	TPK	Passed
Student 2	Competent	Competent	Quite Competent	TPK	Passed
Student 3	Competent	Competent	Competent	TPK	Passed
Student 4	Competent	Not Yet Competent	Not Yet Competent	TPK	Did Not Pass
Student 5	Quite Competent	Quite Competent	Quite Competent	TPK	Passed
Student 6	Not Yet Competent	Competent	Not Yet Competent	TPK	Did Not Pass
Student 7	Quite Competent	Competent	Competent	TEI	Passed
Student 8	Competent	Quite Competent	Quite Competent	TEI	Passed
Student 9	Competent	Competent	Competent	TEI	Passed
Student 10	Competent	Competent	Competent	TEI	Passed
Student 11	Quite Competent	Not Yet Competent	Not Yet Competent	TEI	Did Not Pass
Student 12	Not Yet Competent	Quite Competent	Not Yet Competent	TEI	Did Not Pass
Student 13	Competent	Competent	Quite Competent	LPS	Passed
Student 14	Competent	Competent	Competent	LPS	Passed
Student 15	Quite Competent	Quite Competent	Quite Competent	LPS	Passed
Student 16	Quite Competent	Competent	Quite Competent	LPS	Passed
Student 17	Competent	Not Yet Competent	Not Yet Competent	LPS	Did Not Pass
Student 18	Not Vot		Not Yet Competent	LPS	Did Not Pass

Data Sample Analysis

Upon acquiring the prospective student data, the qualities are delineated and computed according to the predetermined parameters to ascertain the entropy and gain values. The property with the maximum gain value is chosen

as the root. The subsequent table presents the computations of entropy and gain values for each variable, derived from the data in Table 1, utilizing the C4.5 algorithm for the prospective students of Vocational High School Piramida for the 2024/2025 academic year:

Table 2. Entropy and Gain

		Number of Cases (S)	Passed (S1)	Failed (S2)	Entropy	Gain
Total		133	80	53	0,970064152	
Aptitude Test						0,516054161
	competent	53	44	9	0,657272978	
	quite competent	51	36	15	0,873981048	

		Number of Cases (S)	Passed (S1)	Failed (S2)	Entropy	Gain
	not yet competent	29	0	29	0	
Psychometric Test						0,426514729
	competent	51	36	15	0,873981048	
	quite competent	67	44	23	0,927926262	
	not yet competent	15	0	15	0	
Numerical Test						0,54157362
	competent	36	27	9	0,811278124	
	quite competent	64	53	11	0,661976064	
	not yet competent	33	0	33	0	
Major Interests						0,390806677
	TPK	49	32	17	0,931304369	
	TEI	45	25	20	0,99107606	
	LPS	39	23	16	0,976634911	

The Total Entropy column row in the table above is calculated using the following entropy formula equation:

$$\mathit{Entropy}\left(\mathcal{S}\right) = \sum_{i=1}^{n} -p_{i} * \log_{2}\left(p_{i}\right)$$

$$Entropy \; (Total) = \left(-\frac{80}{133}*log_2\left(\frac{80}{133}\right)\right) + \left(-\frac{53}{133}*log_2\left(\frac{53}{133}\right)\right)$$

Entropy(Total) = 0,970064152

Meanwhile, the Gain value in the Interest and Talent Test row is calculated using the following gain formula equation:

$$\begin{aligned} & Gatn\left(S,A\right) = Entropy\left(S\right) - \sum_{i=1}^{m} \frac{|S_{\ell}|}{|S|} * Entropy\left(S_{\ell}\right) \\ & \text{Gain (Potal, fire, short)} - \frac{\tilde{\gamma}_{\ell}}{|S|} \text{(Potal)} + \frac{1}{|S|} * Entropy\left(\text{(ire, short)}, \text{label}\right) \end{aligned}$$

$$Galn\left(Total, tex_minat_bakat \right) \\ = 0,970064152 - \left(\frac{53}{133} * 0,657272978 \right) + \left(\frac{51}{133} * 0,873981048 \right) + \left(\frac{29}{133} * 0 \right) \\ = 0.970064152 - \left(\frac{53}{133} * 0,657272978 \right) + \left(\frac{51}{133} * 0,873981048 \right) + \left(\frac{29}{133} * 0,657272978 \right) + \left(\frac{51}{133} * 0,873981048 \right) + \left(\frac{29}{133} * 0,657272978 \right) + \left(\frac{51}{133} * 0,873981048 \right) + \left(\frac{29}{133} * 0,657272978 \right) + \left(\frac{51}{133} * 0,657272978 \right) + \left(\frac{51$$

Gain (Total,tes_minat_bakat) = 0,516054161

The table above indicates that the numeric test variable will serve as the initial node in the decision group due to its superior gain value.

Subsequently, the second node is chosen, and this process continues until an identical calculation is

employed to derive the tree selection. The

resultant decision tree is as follows:

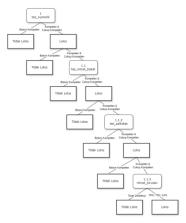


Figure 1. The New Student Admission

Classification Decision Tree

The decision tree results indicate that the quantitative test variable is the most significant factor affecting the graduation of prospective students. This variable serves as the root node of the decision. Consequently, if prospective PPDB applicants achieve elevated numeric exam scores, they will qualify. The scores from the Psychological Test, Aptitude Test, and Major Interest Test are not taken into account.

Implementation of Website Display

The implementation, utilizing the PHP programming language and the Laravel

framework, has yielded a user-friendly the New Student Admission web page for enrolling pupils aspiring to attend Vocational High School Piramida Rancaekek. This system has been developed to facilitate classification functionalities utilizing the C4.5 Algorithm for the New Student Admission committee and personnel. This facilitates a more equitable approach to entering prospective student data and generating reports.



Figure 2. Account Registration View



Figure 3. Landing Page View



Figure 4. Account Login View



Figure 5. Applicant Dashboard View



Figure 6. PPDB Form Display for Applicants



Figure 7. Registrant Logout View



Figure 8. Confirmation of Registrant Logout



Figure 9. Admin Dashboard View



Figure 10. User Data Display



Figure 11. Display of the Add User Data Form



Figure 12. Student Menu Display

Laporan Data Siswa

No	NIK	NISN	Nama Siswa	Jenis Kelamin	Asal Sekolah	Alamat	Minat Jurusan
1	3201748008512440	0022778124	Puspa Suurtini M.Kom.	P	SMP BPI Rancaekek	Dk. Laswi No. 412, Lubuklinggau 61483, Lampung	Teknik Elektro Industri
2	3207972060805628	0028305557	Vanya Utami	L	SMP Lugina	Rn. Industri No. 348, Padang 25552, Banten	Layanan Perbanka Syariah
3	3207153754610267	0020229144	Maida Wijayanti	P	SMP PGRI Haurpugur	Dk. Baya Kali Bungur No. 346, Tarakan 58748, Sumsel	Teknik Elektro Industri
4	3204776241313353	0025571711	Okto Budiyanto	P	SMP PGRI Haurpagar	Jr. Wahid No. 925, Bendar Lampung 88583, Kaltara	Layanan Perbanka Syariah
5	3208060085654089	0027695621	Anastasia Puspa Winarsih	L	SMP PGRI Haurpugar	Jin. Barat No. 579, Makassar 68900, Pabur	Layanan Perbanka Syariah
6	3208175114317937	0024877998	Satya Sawarno M.Farm	L	SMP PGRI Haurpugur	Gg, Padang No. 222, Yogyakarta 40701, Saltra	Layanan Perbanka Syariah
7	3208705292584618	0028236164	Putu Hakim	L	SMP Pasundan Rancaekek	Dk. Panjaitan No. 399, Samarinda 89715, Kalbar	Teknik Pembuata Kain
8	3200857033699318	0029059202	Gilda Oni Laksmiwati S.E.	L	SMP Bakti Ilham	Ki. Astans Anyar No. 553, Gorontalo 66537, DKI	Teknik Elektro Industri

Figure 13. Display of Student Report Print Results



Figure 14. C4.5 Classification Menu Display



Figure 15. Admin and Officer Logout Display

This is a comparative table assessing the effectiveness of the existing system against the proposed system devised by the author:

Table 3 Comparison of Effectiveness Levels

		Running System		P	roposed System
No	Job description	Y/N	Working Process	Y/N	Working Process
1	Registrationn		60-180 Minute		15 Minute
2	Filling out the Registration Form		30-60 Minute		2-5 Minute
3	The completeness of document		30-60 Minute		2-5 Minute
4	Selection Process		1-2 Sunday		15 Minute

		Running System		P	roposed System
No	Job description	Y/N	Working Process	Y/N	Working Process
5	C4.5 Algorithm Selection Results		-		15 Minute
6	Report		1-4 Sunday		5-15 Minute

CONCLUSION

The system has performed satisfactorily, as anticipated. Testing has verified the efficacy of both input and output, particularly in terms of the precision and effectiveness of the C4.5 classification algorithm. Consequently, this system can aid the school in objectively evaluating students as a decision-making support tool in the new student admission process. Furthermore, the report automation feature improves operational efficiency, expediting and streamlining the new student admission process at Vocational High School Piramida.

REFERENCES

- Abrori, M. A. M., Syukur, A., Affandy, A., & Soeleman, M. A. (2022). Improving C4.5 Algorithm Accuracy With Adaptive Boosting Method For Predicting Students in Obtaining Education Funding. *Journal of Development Research*, 6(2), 137–140. https://doi.org/10.28926/jdr.v6i2.205
- Aji, L. I. P., & Sunyoto, A. (2020). An Implementation of C4.5 Classification Algorithm to Analyze student's Performance. 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 126–130.

 https://doi.org/10.1109/ICOIACT50329.2

020.9332088

Anwar, S. (2025). Application of K-Means and C4.5 Algorithms for dropout risk prediction in vocational high schools. *Indonesian Journal of Multidisciplinary Science*, 4(5), 428–435.
https://doi.org/10.55324/ijoms.v4i5.1102

- Chen, P. (2021). The Application of an Improved
 C4.5 Decision Tree. 2021 7th Annual
 International Conference on Network and
 Information Systems for Computers
 (ICNISC), 392–396.
 https://doi.org/10.1109/ICNISC54316.202
 1.00078
- Daulay, S., Apriani, W., & Perwira, Y. (2022).

 Application Of Data Mining For Prediction
 Of Students Out Of College With The
 Method Algorithm C4.5. Jurnal ICT:
 Information and Communication
 Technologies, 13(1), 1–11.
 https://doi.org/10.35335/jict.v13i1.28
- Lintang, M., Pandiangan, N., & Hyronimus, D. (2022). Use of the C4.5 Algorithm to Analyze Student Interest in Continuing to College. SHS Web of Conferences, 149, 01048. https://doi.org/10.1051/shsconf/20221490 1048
- Muhammad, A. R., & Rahardja, E. (2021).

 Pengaruh Gaya Kepemimpinan
 Transformasional dan Motivasi Kerja
 terhadap Kinerja Karyawan dengan
 Komitmen Organisasional sebagai

Variabel Intervening (Studi pada Karyawan Kantor PT. Pos Indonesia Kudus). *Diponegoro Journal of Management, 10*(1).

- Sari, F. W., & Safii, M. (2021). Application of the C4.5 Algorithm in Teaching Teachers' Skills on Learning Effectiveness. *Journal* of Artificial Intelligence and Engineering Applications (JAIEA), 1(1), 54–59. https://doi.org/10.59934/jaiea.v1i1.54
- Suherman, N. R., Ruuhwan, R., & Sudiarjo, A.

 (2023). Implementation of Data Mining at
 Laboratory Vocational High School Using
 The C4.5 Algorithm to Predict Students
 Major Preferences. Innovation in Research
 of Informatics (Innovatics), 5(2).

 https://doi.org/10.37058/innovatics.v5i2.8
 479
- Sulistiani, H., & Aldino, A. A. (2020).

 DECISION TREE C4.5 ALGORITHM
 FOR TUITION AID GRANT PROGRAM
 CLASSIFICATION (CASE STUDY:
 DEPARTMENT OF INFORMATION
 SYSTEM, UNIVERSITAS
 TEKNOKRAT INDONESIA). Edutic Scientific Journal of Informatics
 Education, 7(1).
 https://doi.org/10.21107/edutic.v7i1.8849
- Tempola, F., Muhammad, M., Maswara, A. K., & Rosihan, R. (2022). Rule Formation Application based on C4.5 Algorithm for Household Electricity Usage Prediction. *Trends in Sciences*, 19(3), 2167. https://doi.org/10.48048/tis.2022.2167
- Yan, Y. (2022). Decision Tree Algorithm in the
 Performance Evaluation of SchoolEnterprise Cooperation for Higher

Vocational Education. *Mathematical Problems in Engineering*, 2022, 1–9. https://doi.org/10.1155/2022/4151168

ORI	GIN	AII.	TV R	FP()RT

SIMILARITY INDEX

INTERNET SOURCES

PUBLICATIONS

STUDENT PAPERS

PRIMARY SOURCES

www.researchgate.net Internet Source

2%

journal.unublitar.ac.id Internet Source

Submitted to UIN Syarif Hidayatullah Jakarta 3 Student Paper

1 %

jurnalfkip.unram.ac.id

1 %

Submitted to mu 5

Internet Source

1 %

Student Paper

4

journal.unnes.ac.id 6 Internet Source

jurnalnasional.ump.ac.id

Dede Rizki Darmawan, Riza Arifudin. 8 "Enhancing Durrotalk Chatbot Accuracy Utilizing a Hybrid Model Based on Recurrent Neural Network (RNN) Algorithm and Decision Tree", JUITA: Jurnal Informatika, 2024

<1%

Publication

Muhammad Rizki, Deni Mahdiana. 9 "Application of Classification Algorithm C 4.5 for Predicting Asset Maintenance", International Journal of Recent Contributions from Engineering, Science & IT (iJES), 2019

<1%

Publication

10	ioinformatic.org Internet Source	<1%
11	Submitted to Colaiste Dhualigh College of Further Education Student Paper	<1%
12	Titin Qowidho, M Zarlis, Erna Budhiarti Nababan, Alfian Agusnady, Berti Sari Sembiring. "Analysis Of c4.5 and ID3 Methods in Determining Student Graduation", Journal of Physics: Conference Series, 2019 Publication	<1%
13	www.polgan.ac.id Internet Source	<1%
14	Yunita Sari Siregar, Boni Oktaviana Sembiring, Hasdiana Hasdiana, Arie Rafika Dewi, Herlina Harahap. "Algortihm C4.5 in mapping the admission patterns of new Students in Engingeering Computer", SinkrOn, 2021 Publication	<1%
15	Dinda Ayu Muthia, Andini Ramadhani, Adi Kurniawan, Raka Irfansyah. "Online Student Admission Application at SMK Al-Basyariah Bojong Gede", SinkrOn, 2019	<1%
16	Sri Lestari, Yulmaini, Aswin, Sulyono, Ruki Rizal Nul Fikri. "Artificial Intelligence-Based System 'SiMoniK' for MSMEs", 2023 IEEE 9th Information Technology International Seminar (ITIS), 2023 Publication	<1%
17	jurnal.unsil.ac.id Internet Source	<1%

Kawuu W. Lin, Ya-Jun Zheng, Ju-Chin Chen, Wei-Chiang Wang, Chao-Chun Chen. "A parallel and distributed C4.5 algorithm in cloud computing environments", Computing, 2025

Publication

Exclude quotes

On On Exclude matches

Off

Exclude bibliography