



Rough Sets In Industry

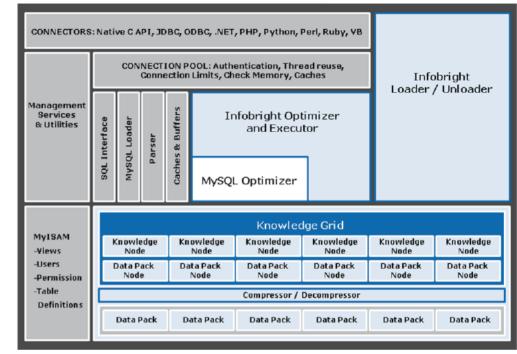
Dominik Ślęzak



My (Industry) Background

Data Analytics 1995-2005
 Data Processing 2005-2015
 Data Science 2015-... D

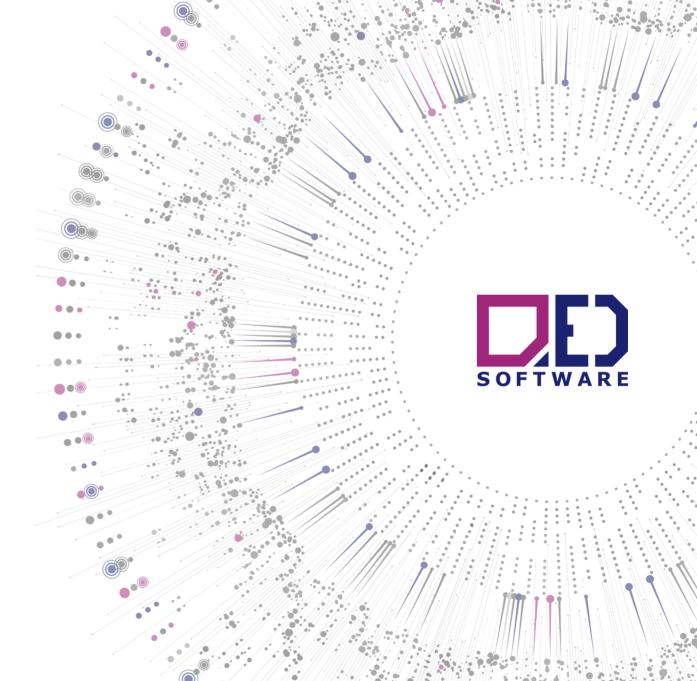
i 🔚 🕒 🎒 🛛 🎼 🖉 🤌 🗎 🔎 🤌 -		J 🎟 🛾 🛉	s 🤹 🛛	Source Sec
		TABLE_SCHEMA	TOTA	RoughSet
ADM_ID	ADMISSIONPRT	HOPITAL	10111	
NODOSS	CRITEREEXCLUSIONPRT	HOPITAL		Explorer
	DIAGPRIPRT	HOPITAL		الارتياب الاتيان المرارك المسلم
- DRG	1 DIAGSECPRT	HOPITAL		
		HOPITAL		
	MEDPRT	HOPITAL	-	
		HOPITAL		
	RESULTATTESTPRT		-	
	HESULIAITESTERT	HOPITAL	-	
ETAT				
- I JRSHOSP			-	
- VARCHARSORT				and the second se
AGE				Contraction of the second second
				A REAL PROPERTY AND A REAL
				PERSONAL PROPERTY (CONTRACT)
				A A ABR II AU
			1000	
- DIAGSECPRT		2		and the second s
B-B DRG_CMDPRT				
DRG		Set 1		
COUM		1	-	
		1 (a a a a a a a a a a a a a a a a a a a	100	
DESCCMD				
- DESCR				and the second s
			-	
				A REAL PROPERTY AND ADDRESS OF AD
- MEDPRT				And a state of the
- DPERATIONPRT				
				Contraction of the local division of the loc
				CARLES AND A COMPANY OF A COMPA
CODEOP -			100	And in case of the local division of the loc
DESCOP				
				and the second se
		and the second second		CONTRACTOR OF THE OWNER.
VARCHARHEURETEST				A REPORT OF THE OWNER OWNER OF THE OWNER OWNE
TYPETEST				and the second s
- B DESCTEST				
NO_MEDPRT			-	
COMMENTS				The second se
RESULTAT			-	



INFOBR GHT

Information Builders Builders

Al for Pioneers



Rough Sets

Zdzisław Pawlak¹

Received June 1981; revised Semptember 1982

We investigate in this paper approximate operations on sets, approximate equality of sets, and approximate inclusion of sets. The presented approach may be considered as an alternative to fuzzy sets theory and tolerance theory. Some applications are outlined.

KEY WORDS: Artificial intelligence; automatic classification; cluster analysis; fuzzy sets; inductive reasoning; learning algorithms; measurement theory; pattern recognition; tolerance theory.

Apart from the known and the unknown, what else is there? Harold Pinter (The Homecoming)

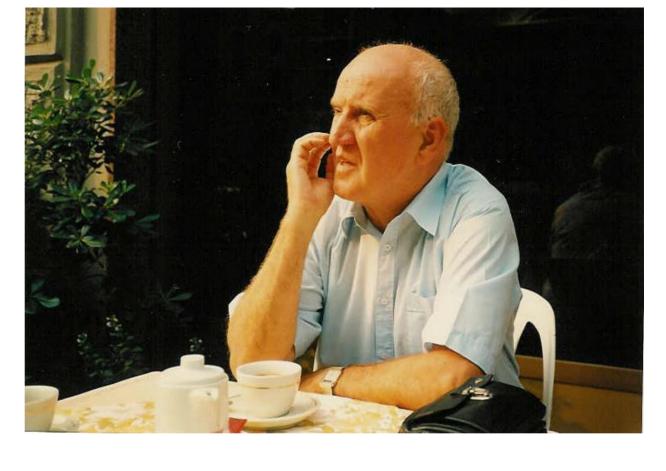
1. INTRODUCTION

The aim of this paper is to describe some properties of rough sets, introduced in Ref. 7 and investigated in Refs. 1, 2, 4, 5, 6, 8, 9, and 11.

The rough set concept can be of some importance, primarily in some branches of artificial intelligence, such as inductive reasoning, automatic classification, pattern recognition, learning algorithms, etc.

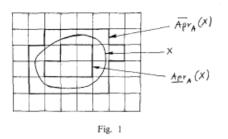
The idea of a rough set could be placed in a more general setting, leading to a fruitful further research and applications in classification theory, cluster analysis, measurement theory, taxonomy, etc.

The key to the presented approach is provided by the exact mathematical formulation of the concept of approximative (rough) equality of sets in a given approximation space; an approximation space is understood as a pair (U, R), where U is a certain set called universe, and $R \subset U \times U$ is an indiscernibility relation. We assume throughout this paper that R is an equivalence relation.



Zdzisław Pawlak 1926-2006 Rough Sets 1982-2022

¹ Institute of Computer Sciences, Polish Academy of Sciences, P.O. Box 22, 00-901 Warsaw, PKiN.



Sets $\operatorname{Edg}_A(X) = X - \operatorname{Apr}_A(X)$ (in short $\operatorname{Edg}(X)$) and $\operatorname{Edg}_A(X) = \overline{\operatorname{Apr}}_A(X) - \overline{X}$, (in short $\operatorname{Edg}(X)$) are referred to as an *internal* and an *external edge* of X in A, respectively.

Of course $\operatorname{Bnd}_4(X) = \overline{\operatorname{Edg}}_4(X) \cup \operatorname{Edg}_4(X)$.

Fig. 1 shows the notion of an upper and lower approximation in a twodimensional approximation space consisting of a rectangle partitioned into elementary squares.

Let us define two membership functions \subseteq_A , $\overline{\in}_A$ (called *strong* and *weak* membership, respectively), as follows:

 $x \in A X \quad \text{iff} \quad x \in \underline{Apr}_{\mathcal{A}}(X)$ $x \in A X \quad \text{iff} \quad x \in \overline{Apr}_{\mathcal{A}}(X)$

If $x \in A$, we say that "X surely belongs to X in A," while $x \in A$ is to mean that "X possibly belongs to X in A." Thus we can interpret approximations as counterparts of necessity and possibility in modal logic.

Of course,

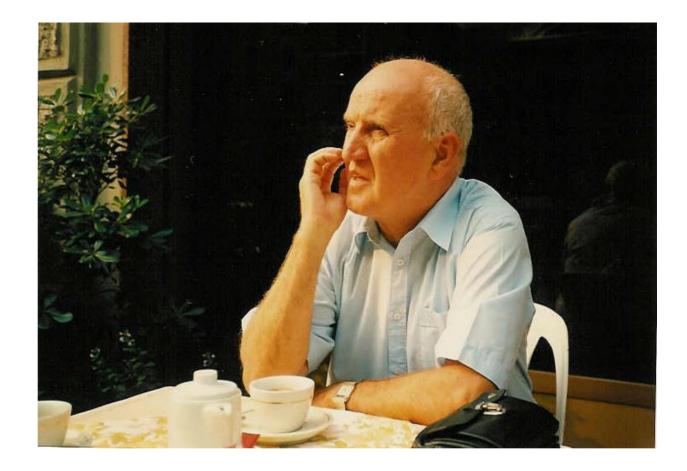
 $\underline{\operatorname{Apr}}_{A}(X) = \{x : x \in X\}$ $\overline{\operatorname{Apr}}_{A}(X) = \{x : x \in X\}$

Thus we can develop our theory in terms of strong and weak membership functions or in terms of approximations. For the sake of simplicity we shall use here the approximational approach.

2.2. Approximation Space and Topological Space

It is easy to check that the approximation space A = (U, R) defines uniquely the topological space T(A) (in short T_A), where $T_A = (U, \text{Com}(A))$, and Com(A) are the family of all open sets in T_A , and U/R is a base for T_A .

It follows from the definition of (lower and upper) approximations that Com(A) is both the set of all open and closed sets in T_A . Thus, $\underline{Apr}_A(X)$ and



Approximations, their calculus and examples

Thus, we can interpret the approximation space $A = (R^+, S)$ as a measurement system, where

 $\bar{\mu}_{A}(i, i+1) = \mu_{A}(i, i+1) = 1, i = 0, 1,...$

is the *unit of measurement* in A, and $\eta(0, r)$ is the accuracy of (0, r) in A. For more detail see Ref. 6.

Example 2. Let V be a finite set called a *vocabulary* and let V^* be the set of all finite sequences over V. Any subset of V^* will be called a *language* over V.

Let $R \subset V^* \times V^*$ be an *indiscernibility* relation, and let $A = (V^*, R)$ be an approximation space defined by V^* and R.

A language $L \subset V^*$ is recognizable in A if $\operatorname{Apr}_A(L) = \overline{\operatorname{Apr}}_A(L)$.

The family of all recognizable languages in \overline{A} , denoted as $\operatorname{Rec}(A)$, is the topology induced by $A = (V^*, R)$ and the base of the topology is V^*/R .

That is to say that if the language L is not recognizable in A we are able to recognize only the lower and upper approximations in A.

This property can be used in speach recognition, pattern recognition, fault tolerant computers, etc.

Example 3. Let $S = \langle X, A, V, \rho \rangle$ be an information system (see Ref. 10), where

X is the set of objects

A is the set of attributes

 $V = \bigcup V_a$, V_a is the set of values of attribute $a \in A$

 $\rho: X \times A \to V$ is an information function, $\rho_X: A \to V$

 $x \in X$ is called an *information about* x in S, where

 $\rho_x(a) = \rho(x, a)$

for every $x \in X$ and $a \in A$.

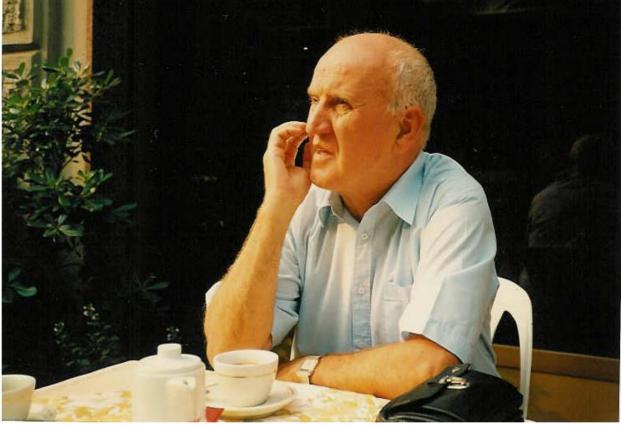
We define the binary relation \tilde{S} over X in the following way:

 $x \sim_S y$ iff $\rho_x = \rho_y$

Obviously \tilde{S} is an equivalence relation and $A = (X, \tilde{S})$ is the approximation space induced by the information system S.

Any subset $Y \subset X$ is called *describable* in S iff $\underline{\operatorname{Apr}}_{A}(Y) = \overline{\operatorname{Apr}}_{A}(Y)$. The set of all describable sets in S, denoted as $\operatorname{Des}(S)$, is a topology induced by S on X, and the base of the topology is X/\tilde{S} .

r of (0, r) in A.



Towards decision reducts (how approximations change when we add / remove attributes)

Transactions on Rough Sets XXII

James F. Peters · Andrzej Skowron Editors-in-Chief



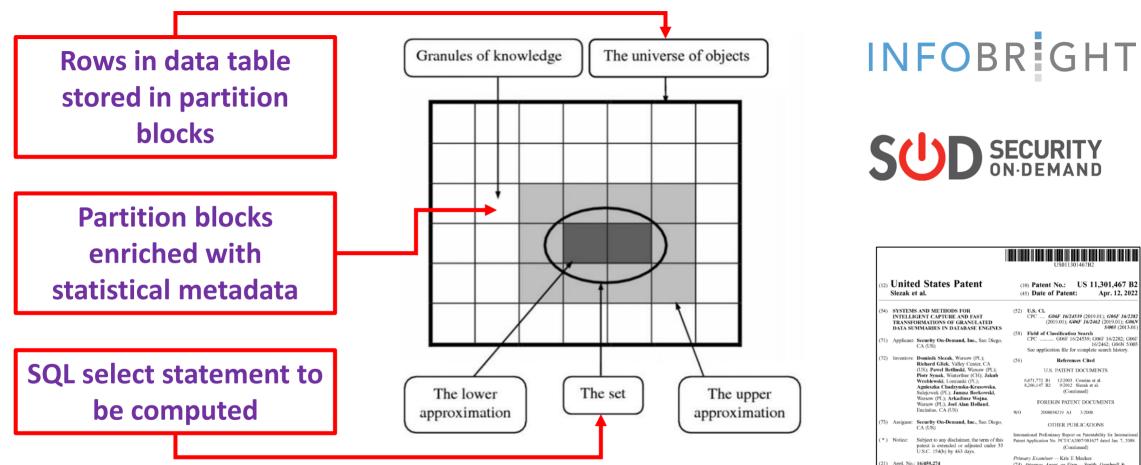
ACM Transactions on Intelligent Systems and Technology Annals of Pure and Applied Logic Applied Intelligence Applied Soft Computing Artificial Intelligence Artificial Intelligence Review **BMC** Bioinformatics Communications of the ACM European Journal of Operational Research Expert Systems with Applications Fundamenta Informaticae Fuzzy Sets and Systems Group Decision and Negotiation IEEE Transactions on Computational Social Systems **IEEE Transactions on Evolutionary Computation** IEEE Transactions on Fuzzy Systems IEEE Transactions on Geoscience and Remote Sensing IEEE Transactions on Image Processing

IEEE Transactions on Knowledge and Data Engineering **IEEE Transactions on Neural Networks** IEEE Transactions on Systems. Man and Cybernetics IEEE/ACM Transactions on Audio, Speech, and Language Processing IEEE/ACM Transactions on Computational Biology and Bioinformatics Information Sciences International Journal of Approximate Reasoning International Journal of Computational Intelligence Systems International Journal of Machine Learning and Cybernetics International Journal of Molecular Science International Journal of Science and Engineering Journal of Applied Non-Classical Logics Journal of Biomedical Informatics **Knowledge and Information Systems** Knowledge Based Systems Sensors Neural Computing and Applications Neural Networks Neural Processing Letters



Neurocomputing Pattern Recognition Pattern Recognition Letters Pharmaceutics Sensors Studia Logica Theoretical Computer Science Web Intelligence and Agent Systems...

RSDS Rough Set Database System					HOME CONTACT				US011301467B2				
HOME SEARCH	SEND STATISTICS	OPINIONS PEOPLE	SOFTWARE MAP	HELP	CONTACT	(12)	Unite Slezak o	ed States Patent et al.	(10) Patent No.: (45) Date of Patent:	US 11,301,467 B2 Apr. 12, 2022			
USER MENU	Page 1 Page		te 5 Page 6 Page 7			(71)	INTELLI TRANSF DATA SU Applicant	S AND METHODS FOR GENT CAPTURE AND FAST ORMATIONS OF GRANULATED MMARIES IN DATABASE ENGINES Security On-Demand, Inc., San Diego, CA (US) Dominik Slezak, Warsaw (PL);	(2019.01); Got (58) Field of Classification CPC	24539; G06F 16/2282; G06F 16/2462; G06N 5/003 complete search history.			
USER LOGIN Usemame:	Number of register Number of authors	en visited 3363048 times. ed users: 405. (in the database): 42859. publications are included	?					Richard Glick, Valley Center, CA (US): Pavel Betlinski, Warasow (PL); Piotr Synak, Winterthur (CH); Jakub Wroblewski, Lomianki (PL); Agniezska Chadzynska-Krasowska, Sulejowić (PL): Janusz Berkowski, Warssw (PL); Jael Alan Helland, Enclinitas, CA (US)	U.S. PATENT I 6,671,772 B1 122003 (8,266,147 B2 9/2012 S (Conti FOREIGN PATEN WO 2008034219 A1	DOCUMENTS Cousins et al. Slezak et al. nued) TI DOCUMENTS			
Password:	Mind of sublicit	on Number of publications		_		(73)	Assignce:	Security On-Demand, Inc., San Diego, CA (US)	OTHER PUB				
	article	on Number of publications		21,552		(*)	Notice:	Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 463 days.	International Preliminary Report Patent Application No. PCU/CA2 (Conti	007/001627 dated Jan. 7, 2008.			
Log in O Create new account	incollection	-		1,232		11° 1		: 16/459,274	Primary Examiner — Kris E 1 (74) Attorney, Agent, or Firm Russell LLP				
Request new password	book			176		(/	Filed:	Jul. 1, 2019	(57) ABSTI	RACT			
Request new password	proceedings inbook			144 60 26		(65)	US 20208	Prior Publication Data 0004749 A1 Jan. 2, 2020	Embodiments may provide m- ligent capture and fast transf summaries. An engine may be summaries into result sets repr	formation of granulated data e used to transform input data			
PATRONS OF SERVICE				20		11	Re	lated U.S. Application Data	data summaries contain enoug	h knowledge about the origi-			
PATRONS OF SERVICE	mastersthesis			12		(60)	Provisiona 29, 2018.	al application No. 62/691,751, filed on Jun.	nal data to accurately perform without needing to access the ment, the contents of data sun	original data. In an embodi-			
🛞 👱 ኲ 🎸	manual	L		2		(61)			SQL approximate engine which	ch retrieves summaries stored			
	software	L		19		(51)	Int. Cl. G06F 16/ G06F 16/	2453 (2019.01)	on disk and utilizes them for its contents of data summaries are which give users direct access	e accessible via virtual tables to the summary contents and			
	Total:	38549						(Continued)	(Conti	nued)			



M Wnuk, S Stawicki, D Ślęzak: Reinventing Infobright's Concept of Rough Calculations on Granulated Tables for the Purpose of Accelerating Modern Data Processing Frameworks. IEEE BigData 2020: 5405-5412

(74) Attorney, Agent, or Firm - Smith, Gambrell & Russell LLP (57) ABSTRACT Embodiments may provide methods and systems for inte ligent capture and fast transformation of granulated da summaries. An engine may be used to transform input dat summaries into result sets representing query outcomes. Th data summaries contain enough knowledge about the orig nal data to accurately perform operations on the summa (60) Provisional application No. 62/691,751, filed on Jun without needing to access the original data. In an embod ment, the contents of data summaries are accessible via a SQL approximate engine which retrieves summaries store on disk and utilizes them for its operations. Alternatively, th contents of data summaries are accessible via virtual tables which give users direct access to the summary contents (Continued)

Jul. 1, 2019

US 2020/0004749 A1 Jan. 2, 2020

(51) Int. CL. G06F 16/2458

G06F 16/2453

Prior Publication Data

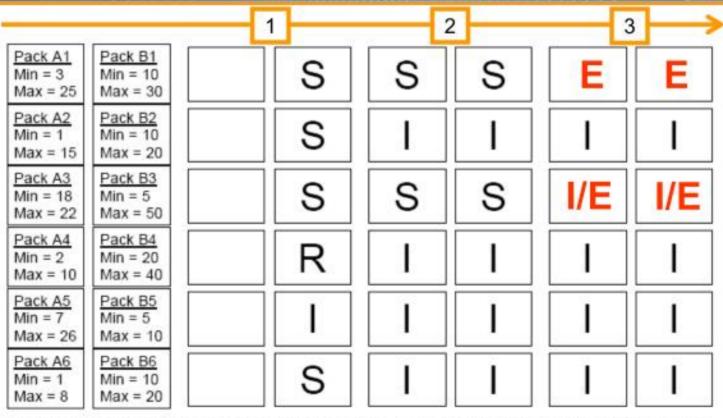
(2019.01)

(2019.01)

Related U.S. Application Data

(Continued)

SELECT MAX(A) FROM T WHERE B>15;

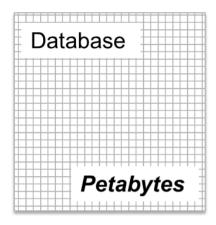


INFOBR GHT

SUD SECURITY ON-DEMAND

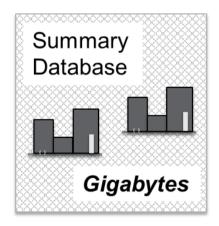
	US011301467B2							
	United States Pat Slezak et al.	ent (10) Patent No.: US 11,301,467 B2 (45) Date of Patent: Apr. 12, 2022						
(54)	SYSTEMS AND METHODS FOR INTELLIGENT CAPTURE AND TRANSFORMATIONS OF GRAM DATA SUMMARIES IN DATAB/	FAST CPC G06F 16/24539 (2019.01); G06F 16/282 NULATED (2019.01); G06F 16/2462 (2019.01); G06N						
(71)	Applicant: Security On-Demand, I CA (US)							
(72)	Inventors: Dominik Stezak, Warsy (US); Pawel Betlinski, Piotr Synak, Winterfluu Wroblewski, Lomianki Agnieszka Chadzynska Sulejowk (PL); Janusz Warsw (PL); Arkadius Warsw (PL); Joel Alan Encinitas, CA (US)	Content CA (50) Reservances Warnaw (PL) U.S. PATINT DOCUMENTS (C11), Jakub 6671.772 B1 12/2003 (C10), Constance, a S266.147 B2 2/2012 Stack et al. Derkrowski, (Continued) 2/2012						
(73)	Assignce: Security On-Demand, I CA (US)	OTHER PUBLICATIONS						
(*)	Notice: Subject to any disclaimen patent is extended or ac U.S.C. 154(b) by 463 da	ljusted under 35 (Continued) nys.						
(21)	Appl. No.: 16/459,274	Primary Examiner — Kris E Mackes (74) Attorney, Agent, or Firm — Smith, Gambrell & Russell LLP						
<u>`</u>	Filed: Jul. 1, 2019	(57) ABSTRACT						
(65)	Prior Publication Da US 2020/0004749 A1 Jan. 2, 20 Related U.S. Application D	Embodaments may provide memods and systems for fine- 20 ligent capture and fast transformation of granulated data summaries. An engine may be used to transform input data summaries into result sets representing query outcomes. The data summaries contain enough knowledge about the origi- nal data to accurately ereform operations on the summaries						
	Provisional application No. 62/691,7 29, 2018. Int. Cl.	751, filed on Jun. without needing to access the original data. In an embodi- ment, the contents of data summaries are accessible via ar SQL approximate engine which retrieves summaries store on disk and utilizes them for its operations. Alternatively, the						
(01)	G06F 16/2458 (2019.01) G06F 16/2453 (2019.01) (Continued)	on disk and unizes them for its operations. Alternatively, the contents of data summaries are accessible via virtual tables which give users direct access to the summary contents and (Continued)						

I/S/R denotes irrelevant/suspect/relevant; E - exact computation (decompression)



Traditional Query Execution:

- long time to do computations
- lots of disk/memory/processing resources required
- hard to manage in data lake
 / data cloud environments

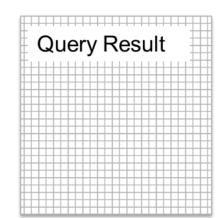


Querying on Data Summaries:

- orders of magnitude faster (original operations replaced by fast summary transformations)
- far less resources consumedoriginal data remaining in-place

M Bartoszuk, J Litwin, M Wnuk, D Ślęzak:

Tensor-based Approach to Big Data Processing and Machine Learning. IEEE BigData 2022



Approximate Query Result (accurate enough from business perspective)





	US011301467B2									
(12)	Unite Slezak e	d States Patent t al.	()) Patent					67 B2 2, 2022	
(54)	INTELLI TRANSF	S AND METHODS FOR GENT CAPTURE AND FAST DRMATIONS OF GRANULATED MMARIES IN DATABASE ENGINES			19.01); G	06F 16/.	2462 (2019.0	F 16/2282 01); G06N (2013.01	
(71)	Applicant:	Security On-Demand, Inc., San Diego, CA (US)		Field of Cla CPC See applicat	G06F 10	5/24539;	G06F 16/24	62; G	06N 5/00	
(72)	Inventors:	Dominik Slezak, Warssw (PL); Richard Glick, Valley Center, CA (US); Pawel Betlinski, Warsaw (PL); Piotr Synak, Winterhur (PH); Jakub Wroblewski, Lomianik (PL); Agnieszka Chadzynska-Krasowska, Salejowek (PL); Janusz Borkowski, Warsaw (PL); Jackadiusz Wojna, Warsaw (PL); Joel Alan Holland, Encinitias, CA (US)		671,772 B1 266,147 B2 FOREI	9/2012	DOCU Cousins Slezak o tinued)	MEN et al. et al. CUME			
(73)	Assignce:	Security On-Demand, Inc., San Diego, CA (US)		O	THER PU	BLICAI	TIONS	;		
(*)	Notice:	Subject to any disclaimer, the term of thi patent is extended or adjusted under 3: U.S.C. 154(b) by 463 days.	s Patent /	International Preliminary Report on Patentability for Internation Patent Application No. PCT/CA2007/001627 dated Jan. 7, 2008. (Continued)						
(21)	Appl. No.: 16/459,274			Primary Examiner — Kris E Mackes (74) Attorney, Agent, or Firm — Smith, Gambrell & Russell LLP						
(22)	Filed:	Jul. 1, 2019	(57)		ABS	PACT				
(65)) Prior Publication Data US 2020/0004749 A1 Jan. 2, 2020			Embodiments may provide methods and systems for intel ligent capture and fast transformation of granulated data summaries. An engine may be used to transform input data summaries into result sets representing query outcomes. The						
Related U.S. Application Data			nal dat	data summaries contain enough knowledge about the origi- nal data to accurately perform operations on the summaries						
(60)	29, 2018. Int. Cl. G06F 16/2		 withou ment, t SQL ap on disk content 	t needing to the contents pproximate and utilizes ts of data su	access the of data su engine whether for i mmaries a	e origin immarie ich retri its opera ire acces	al dat s are a eves s tions ssible	a. In a accessi aumma Alterna via vir	n embodi ible via ar ries stored atively, the tual table	
	G06F 16/2		which	give users d			summ	nary co	ntents and	
		(Continued)			(Con	tinued)				