Welcome To DCMI Special Session: Applying Taxonomies in Publishing – Leveraging Your Semantic Enrichment Investment

13 October 2016, 10:30 to 12:00

**Access Innovations, Inc.** 

Changing search to found

## Jay Ven Eman, Ph.D., CEO

j ven eman@accessinn.com www.accessinn.com www.dataharmony.com +1.505.998.0800 Albuquerque, NM



Publishers ask, "I have a taxonomy! Now what?"

- What do you plan to do with your taxonomy?
  - Print it and...
    - Put your monitor on top of it
    - Use it as a door stop
    - Send it to Dad along with your student loan
  - Archive it and...
    - Send a copy to your boss
    - Post it on FaceBook
    - Turn it into a Broadway musical
    - Make a documentary and put it on YouTube



# What do you have in mind?

- Your ideas...
- Who has a
  - Taxonomy, thesaurus, controlled vocabulary, ontology, etc?
- Who plans on getting one?
- Why?



# Use Case 1 - Maintain your thesaurus

- Who someone must be tasked with the responsibility
- How many a part time person to two FTEs
- Process
  - Mine contributions
    - Subject matter experts (SME)
    - Authors
    - Users
    - Search logs



# Use Case 1 - Maintain your thesaurus

## Process cont.

- Review source material
- Tools for maintenance
- How often
  - Analyze frequency of new terms and changes to existing terms
  - Set a fixed schedule





# **Content Publishing Pipeline**



## Use Case 2 – Index your content



# **Content Publishing Pipeline**





Data Harmony MAIstro Server

# **Content Publishing Pipeline**



## Use Case 2 – Index your content



# Sample unstructured document

 Next
 Zoom
 Move
 Text
 Select
 Annotate
 Sidebar
 Search

 3790
 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 54, NO. 8, AUGUST 2008

#### Constructions of Multiblock Space–Time Coding Schemes That Achieve the Diversity–Multiplexing Tradeoff

Previous

Hsiao-Feng (Francis) Lu, Member, IEEE

Abstract—Constructions of multiblock space-time coding schemes that are optimal with respect to diversity-multiplexing (D-M) tradeoff when coding is applied over any number of fading blocks are presented in this correspondence. The constructions are based on a left-regular representation of elements in some cyclic division algebra. In particular, the main construction applies to the case when the quasi-static fading interval equals the number of transmit antennas, hence the resulting scheme is termed a minimal delay multiblock space-time coding scheme. Constructions corresponding to the cases of nonminimal delay are also provided. As the number of coded blocks approaches infinity, coding schemes derived from the proposed constructions can be used to provide a reliable multiple-input multiple-output (MIMO) communication with vanishing error probability.

Index Terms—Cyclic-division algebras, diversity-multiplexing (D-M) tradeoff, fading channels, multiplock space-time codes, multiple-input multiple-output (MIMO) channels, number fields, space-time codes.

#### I. INTRODUCTION

By deploying multiple antennas at both transmitter and receiver ends, the multiple–input multiple–output (MIMO) technology can significantly increase the ergodic channel capacity as well as improve the link reliability. For example, in a MIMO communication system with  $n_t$  transmit and  $n_r$  receive antennas, under the quasi-static MIMO Rayleigh block fading channel model, it is known [1] that the ergodic MIMO channel capacity C equals

$$C = \min\{n_t, n_r\} \log_2 \text{SNR} + O(1) \text{ bits/channel use}$$
(1)

#### at high signal-to-noise ratio (SNR) regime.

Coding schemes dedicated to the MIMO systems to achieve higher transmission rate and better link reliability are specifically coined

the multiblock code  $\mathcal{X}$  with multiplexing gain r, we say  $\mathcal{X}$  achieves *diversity gain* d(r) if at high SNR regime, the codeword error probability of  $\mathcal{X}$  is on the order of

$$P_e(r) \doteq SNR^{-d(r)}$$
. (4)

By  $\doteq$ , we mean the exponential equality defined in [4]. We say the function  $f(SNR) \doteq SNR^b$  if and only if

$$\lim_{SNR\to\infty} \frac{\log f(SNR)}{\log SNR} = b.$$
 (5)

The notations of  $\geq$  and  $\leq$  are defined similarly.

In their ground-breaking paper, Zheng and Tse [4] showed that there exists a fundamental tradeoff between the multiplexing (D-M) tradeoff. For the cases when  $T \ge n_t + n_r - 1$  and when the *m* consecutive fading blocks are statistically independent, the D-M tradeoff asserts that the maximum possible diversity gain  $d^*(r)$  for any  $(n_t \times mT)$  multiplex space-time coding schemes with multiplexing gain *r* is a piecewise linear function connecting the points  $(k, d^*(k)), k = 0, 1, \dots, \min\{n_t, n_r\}$ , and

$$d^{*}(k) = m(n_{t} - k)(n_{r} - k).$$
 (6)

On the other hand, if  $T < n_t + n_r - 1$ , only upper and lower bounds on  $d^*(r)$  are available in [4].

This remarkable result has spurred a considerable amount of research activities on constructing space-time coding schemes to achieve this optimal tradeoff  $d^*(r)$ . Much progress has been made in the case of m = 1. When  $n_t = n_r = T = 2$  and m = 1, several D-M optimal  $(2 \times 2)$  schemes with tilted-QAM constellations can be found in [5]-[8]. These are the first instances showing that the D-M tradeoff (6) holds even in the case of  $T < n_t + n_r - 1$ . In particular, the "Golden code" proposed by Belfiore *et al.* [8] is shown to have the best error performance among the aforementioned  $(2 \times 2)$  codes. By generalizing the Golden code construction and working on latices with written comparison.



# Run the documents through Metadata Extractor<sup>™</sup> to create well-formed, rich XML

```
-<Metadata>
   <DocumentName>04567587.pdf</DocumentName>
 -<Title>
    Constructions of Multiblock Space-Time Coding Schemes That Achieve the Diversity-Multiplexing Tradeoff
   </Title>
 -<Publication>
    <Source>IEEE TRANSACTIONS ON INFORMATION THEORY</Source>
     <Volume>54</Volume>
     <Number>8</Number>
   -<PublicationDate>
       <Month>AUGUST</Month>
      <Year>2008</Year>
                                                                  Automatic (per doc template)
    </PublicationDate>
                                                                   E.g. Dublin Core Metadata, JATS
   </Publication>
 -<Page>
                                                                   Bibliographic citation
    <StartPage>3790</StartPage>
    <EndPage>3796</EndPage>
   </Page>
 -<Abstract>
   -<P>
```

Constructions of multiblock space-time coding schemes that are optimal with respect to diversity-multiplexing (D-M) tradeoff when coding is applied over any number of fading blocks are presented in this correspondence. The constructions are based on a left-regular representation of elements in some cyclic division algebra. In particular, the main construction applies to the case when the quasi-static fading interval equals the number of transmit antennas, hence the resulting scheme is termed a minimal delay multiblock space-time coding scheme. Constructions corre-sponding to the cases of nonminimal delay are also



# Automatically add the taxonomy terms

- <subject></subject>	rt>
- <authorpickedterms></authorpickedterms>	horPickedTerms>
<term>Cyclic-division algebras</term>	erm>Cyclic-division algebras
<term>diversity-multiplexing (D-M) tradeoff</term>	erm>diversity-multiplexing (D-M) tradeoff
<term>fading channels</term>	erm>fading channels
<term>multiblock space-time codes</term>	erm>multiblock space-time codes
<term>multiple-input multiple-output (MIMO) channels</term>	erm>multiple-input multiple-output (MIMO) channels
<term>number fields</term>	erm>number fields
<term>space-time codes</term>	erm>space-time codes
	thorPickedTerms>
- <maisuggestedterms></maisuggestedterms>	JSuggestedTerms>
<term>Construction industry</term> Entity extraction: People,	erm>Construction industry
<1erm>Encoding 1erm	erm>Encoding
<term>Fading</term> Places, Things	
<term>Delay</term> Conceptual indexing: using the	Yerm>Delay Conceptual indexing: using the
<term>Transmitting antennas</term> taxonomy	
<term>MIMO</term>	'erm>MIMO
- <authors></authors>	
- <author></author>	
- <name></name>	
<lastname>Lu</lastname>	
<firstname>Hsiao-Feng (Francis)</firstname>	<firstname>Hsiao-Feng (Francis)</firstname>

Innovations Inc.\*

maker of Data Harmony taxonomy software

# Use Case 3 – Integrate





# Use Case 4 – Website navigation



SVSCOM 1

Syscom Services, Inc. Project Manager: Stephanie Connolly Tel: 301-768-0112 | Email: NOTE: This wireframe is sized to emulate an 1024 x 768 pixel web page.

DISCLAIMER : The layouts displayed in this diagram represent information structure only and should not be construed as final designs and/or layouts



SYSCOM /

Tel: 301-768-

R : The layouts displayed in this diagram represent information structure only and should ued as final designs and/or lavouts



# 1,200

Professional Engineers board certified by ASCE, recognizing advanced knowledge in specialty areas of engineering.

#### **GET CERTIFIED**

#### **FIND A GROUP**

USA	INTERNATIONAL					
Enter Zip (	Code	GO				
OR BROWSE ALL GROUPS						

At nearly 1 3/4 miles up, Italy's Stelvio Pass is the highest paved mountain pass in the Eastern Alps. Car enthusiasts consider the 1985 ASCE civil engineering landmark one of the world's greatest roads to drive.



## Use Case 5 – Search Harmony Recommender



Content Management



With Access Innovation's *Search Harmony* navigation layer, typing in the text box prompts the user to select from thesaurus terms for a search.





## Use Case 5 – Seach Harmony Recommender

National Information Center for Educational Media

music theory Search

Results for music theory click here for a key word search

<u>Music Theory Fundamentals</u> Includes ten study charts featuring **music** alphabet, sharps, flats, rhythm, time signatures, key Descriptors : Meter and rhythm - **Music theory** and composition http://www.mediasleuth.com/revmai/bibrec.php?accnum=82275

Music Theory Fundamentals Includes ten study charts featuring music alphabet, sharps, flats, rhythm, time signatures, key Descriptors : Meter and rhythm - Music theory and composition http://www.mediasleuth.com/revmai/bibrec.php?accnum=179810

Practical Music Theory Series

Presents the complete three - part series on practical **music theory** created by Sandy Feldstein Descriptors : Musicology - Musical notation - **Music - Music theory** and composition - Meter and rhythm - Harmony http://www.mediasleuth.com/revmai/bibrec.php?accnum=404250

Essentials of Music Theory - Compact Disc 1

Introduces musical concepts through lessons followed by exercises. Covers note values, time signatures, triads, intervals, scales, and ear training. Includes rhythmic, melc performed by a number of different instruments including piano, flute, clarinet, and cello. Descriptors : **Music theory** and composition - **Music** appreciation http://www.mediasleuth.com/revmai/bibrec.php?accnum=473483

Essentials of Music Theory - Compact Disc 2

Introduces musical concepts through lessons followed by exercises. Covers note values, time signatures, triads, intervals, scales, and ear training. Includes rhythmic, melo performed by a number of different instruments including alto sax, trumpet, trombone, and violin. Descriptors : **Music theory** and composition - **Music** appreciation http://www.mediasleuth.com/revmai/bibrec.php?accnum=473485

A list of content containing the selected thesaurus term is returned. Selecting an item returns its full record...





## Use Case 5 – Search Harmony Recommender

#### National Information Center for Educational Media

### Practical Music Theory Series

Presents the complete three - part series on practical music theory created by Sandy Feldstein for Alfred's Publishing. Covers staff, clefs, measures, time signatures, notes, rests, ties and slurs, repeats and endings, flats, sharps and naturals, scales, key signatures, dynamics, tempos, chords, inversions and harmonizing. Apple version compatible with Commodore computers, IBM version available in 3.5 and 5.25 diskettes. Includes theory books. Descriptors : Musicology - Musical notation - Music - Music theory and composition - Meter and rhythm - Harmony

#### More like this documents:

- Practical Music Theory Volume 3 ; and 5.25 diskettes. Part 3 of a three part series...
- Alfred Music Achievement Series : and rhythmic notation , music terminology and ties...
- Practical Music Theory Volume 2 : on practical music theory created by Sandy Feldstein...
- Practical Music Theory Volume 1 : series on practical music theory created by Sandy...
- Basic Piano Theory Software Levels 1a, 1b, 2 : Alfred's Basic Piano Theory Levels 1 5 and adult ....
- Basic Piano Theory Software Levels 3, 4, 5 : Alfred's Basic Piano Theory Levels 1 5 and adult ....
- Basic Piano Theory Software Adult Level : Alfred's Basic Piano Theory Levels 1 5 and adult ....
- <u>Basic Piano Theory Software Student Version Level 1A</u>: Offers student diskette only, Level 1A, from Alfred's...
- <u>Basic Piano Theory Software Student Version Level 1B</u> : Offers student diskette only, Level 1B, from Alfred's...
- Basic Piano Theory Software Student Version Level 2: Offers student diskette only, Level 2, from Alfred's...

### Thesaurus terms

### Similar content

...as well as suggesting other content with similar semantic fingerprints—not just the original search term.





# Use Case 6 - Related-term/concept navigation – semantic navigation



Navigate the full taxonomy "tree"

## BROWSE

#### Provide a solution (131) biological succession (8) biophysics (1) biorhythms (3) botany (518) cellular biology (153)

Biosynthesis (2) ➡ Cells (Biology) (9) Enzyme production (5) ➡ Nucleic acids (0)

#### Tells the story of the latest and most far - reaching scientific revolution, the growth of bioindu Explores the new techniques and how they will be used. Descriptors : Microbiology - Genetics - **Cellular biology** http://www.mediasleuth.com/revmai/bibrec.php?accnum=1

#### The Building Blocks of Life

Provides a comprehensive look at the basic unit of life, the cell, and how cells function both together and alone. Descriptors : Cellular biology - Cells (Human) http://www.mediasleuth.com/revmai/bibrec.php?accnum=1



000	PLOS ONE : accelerating the publication of peer-reviewed science									
	▶ 🖄 + 🚱 www.plosone.org							C Read	der 🛛 🔘	
C III Access Innova	ations T Apple Apple Bing Facebook (7) Freeki	bblekat Google Goog	le Maps News ▼ Popular	v ▼ SugarC	RM Tools and	Reference 🔻 Travel 🔻	Yahoo!			
PL	PLOS   Public Library Of Science	PLOS ONE : acceleratin	g the publication of peer-rev	iewed scien	ce		EDRM			∫ + ∫ III
				plos	org create account	sign	in			
		Subject Are	ubject Areas For Authors		ıt Us	Search	(	م		
	•					а	dvanced s	earch		
	Browse Subject Areas: Biology and life sciences / Anatomy / Lymphatic system									
	View All Articles (88368) View All Articles		I Articles (86086)	Articles (86086) Histold						
	Biology and life sciences	Agricul	ture	Integumentary system		tary system				
	Computer and information sciences	Anaton	ıy	•	Liver			>		
	Earth sciences	Astrobi ▶	ology	•	Lymphatic	system	Þ			
	Feelenv and environmental		ustics (201)		Musculos	celetal system	•			

In the News Recent Most Viewed

## Breadcrumb searching with a twist





### Egr3 Induces a Th17 Response by Promoting the Development of $\gamma\delta$ T Cells

Rose M. Parkinson, Samuel L. Collins, Maureen R. Horton, Jonathan D. Powell

Published: January 24, 2014 • DOI: 10.1371/journal.pone.0087265



Video

#### Figures

# Use Case 7 - Facilitate eCommerce

- Term disambiguation
  - Sofas couches
  - Motorcycle helmets brain buckets
  - Bikes
    - Motorcycles
    - Tricycles
    - Bicycles
    - Unicycles
- Related terms
  - Bicycle helmets bicycle seats
  - Motorcycles health insurance



# Use Case 8 - Facilitate cross-marketing



SYSCOM 1

Tel: 301-768-

R : The layouts displayed in this diagram represent information structure only and should ued as final designs and/or lavouts

# Use Case 8 - Facilitate cross-marketing



## Use Case 9 – Author networks





### Data Harmony MAIstro





AFF#	Institution name	Department	State	Count		
AF0002649	Royal Melbourne Institute of Technology (RMIT)	Melbourne and the melbourne centre for nanofabrication		Austr	Search f	
AF0007933	ArcelorMittal	NDE - Instrumentation - Control Dept.		france	Affliatior	ns
AF0009777	Farmitalia Carlo Erba-Erbamont, Research Lab, Milano, Italy					
AE0039842	ArcelorMittal Global R&D Gent					_



# Use Case 10 - Semantic fingerprint

- Semantic profiles
  - Your content
  - Members
  - Customers
  - Authors/contributors
  - Companies, institutions
- Personalization, customization, targeting
  - Topics
  - Communities
  - Centers of excellence



# Use Case 10 - Semantic fingerprint

## Author Profile

## Filters by topic, date, publication, co-author, etc.





# Use Case 11 – Find-an-expert, peer review

- Index at manuscript submission
- Map to semantic profiles
  - Authors
  - Members
  - Reviewers



## Use Case 12 - Matchups

- People to people
- People to resources and activities
- People to organizations
- Resources to resources

# Semantics adds precision to the matching process



## Use Case 13 - Mapping terminology differences

- Department to department
  - Finance to marketing
  - Legal to production to...
- Professional to lay person
  - Iron pyrite fool's gold
  - Dionaea muscipula venus flytrap
- Scientific to generic to trade name
  - N-acetyl-p-aminophenol paracetamol acetaminophen Tylenol – Panadol
  - Acetylsalicylic acid aspirin Bayer Aspirin
- Old to new
  - Consumption tuberculosis
  - Light amplification by stimulated emission of radiation LASER
     laser





# Use Case 14 - Different views

## Thesaurus Master<sup>®</sup>

- Hierarchical
- Alphabetical
- All Term Records
- Candidate Terms
- Frequency
- Permutated (Comma)
- Permutated (Tab)
- Deleted Terms
- Dynamic View





## Use Case 15 – Concept/term mining & visualization

## **Relevant Disciplines for Science Mapping**



ACCESS Innovations Inc. maker of Data Harmony taxonomy software



Innovations Inc.\*

maker of Data Harmony taxonomy software

# Use 16 – And more!!

- Image indexing
- Fate predicting & SciGen bogus papers detection
- Mining for grants & funding
- Patent mining & analysis
- Hot-topic page production
- Ad serving
- Analytics for product development, competitive analysis, metrics
- And more!!



# Thank you! Questions?

# Access Innovations, Inc. Changing search to found

Jay Ven Eman, Ph.D., CEO

<u>i ven eman@accessinn.com</u> <u>www.accessinn.com</u> <u>www.dataharmony.com</u> +1.505.998.0800 Albuguergue, NM

