COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCI	EMENT/SOLICITATION	NO./CLOS	SING DATE/if r	ot in response to a pro	ogram announcement/solicit	ation enter NSF 07-140	FOR NSF USE ONLY			
							NSF PF	ROPOSAL NUMBER		
IIS - INFO INT	EGRATION &	INFOR	MATICS							
DATE RECEIVED	NUMBER OF C	OPIES	DIVISION	ASSIGNED	FUND CODE	DUNS# (Data Univers	al Numbering System)	FILE LOCATION		
						009094012	4012			
EMPLOYER IDENTIFICATION NUMBER (EIN) OR SH			IOW PREVIO	US AWARD NO.	IF THIS IS	IS THIS PROPOS	AL BEING SUBMITT	ING SUBMITTED TO ANOTHER FEDERAL		
TAXPAYER IDENTIFIC	ATION NUMBER (TIN)		A RENEWAL	USHMENT-BASE	D RENEWAL	AGENCY? YES	□ NO □ IF YES, LIST ACRONYM(S)			
			414644							
NAME OF ORGANIZAT	ION TO WHICH AWAR		BE MADE	ADDRE						
Brichem Veung Uni			Brigham Young University							
		\ \	A-285 ASB							
)		Provo, UT. 846021231						
0036707000										
NAME OF PERFORMIN	IG ORGANIZATION, IF	DIFFEREN	IT FROM ABC	DVE ADDRES	SS OF PERFORMING	ORGANIZATION, IF L	DIFFERENT, INCLU	DING 9 DIGIT ZIP CODE		
PERFORMING ORGAN	IZATION CODE (IF KNO	OWN)								
						1				
IS AWARDEE ORGANIZATION (Check All That Apply) SMALL BUSINESS MINORITY BUSINESS IF THIS IS A PRELIMINARY PROPOSAL (See GPG II.C For Definitions) FOR-PROFIT ORGANIZATION WOMAN-OWNED BUSINESS THEN CHECK HERE										
TITLE OF PROPOSED	PROJECT Collabo	rative R	esearch: 7	TANGO: Tal	ble Analysis for	Semiautomatic				
	Generat	tion of O	Intologies							
	·				REQUESTED STAR					
\$ 12,000	1	0	months	(1-00 MONTHS)	REQUESTED STAR		APPLICABLE			
	CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW							nce Number		
	OBBYING ACTIVITIES	(GPG II.C)			Exemption Subsec	tion or IRB A	pp. Date			
PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d) INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED										
☐ HISTORIC PLACES	(GPG II.C.2.j)				(GPG II.C.2.j)					
SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)										
UVERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR										
Computer Science										
801-422-0169 Provo, UT United Stat				UI 84002 States						
NAMES (TYPED)		High De	egree	Yr of Degree	25 Degree Telephone Number		Electronic Mail Address			
PI/PD NAME										
David W Embley Pl		PhD		1976	801-422-6470 emb		oley@cs.byu.edu			
CO-PI/PD		_					U			
CO-PI/PD										
CO-PI/PD										
CO-PI/PD										

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the Authorized Organizational Representative or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), nondiscrimination, and flood hazard insurance (when applicable) as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG) (NSF 07-140). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

Conflict of Interest Certification

In addition, if the applicant institution employs more than fifty persons, by electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.A; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be dislosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

Debarment and Suspension Certification (If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

Certification Regarding Lobbying

The following certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding \$100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding \$150,000.

Yes 🗖

No 🗖

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

Certification Regarding Nondiscrimination

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

community in which that area is located participates in the national flood insurance program; and
 building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:

(1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and

(2) for other NSF Grants when more than \$25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

AUTHORIZED ORGANIZATIONAL REP	SIGNATURE	DATE					
NAME							
TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX N		NUMBER			
*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN							
INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.							

This is a request for an REU Supplement for NSF Grant #0414644 for 2 undergraduate students.

REQUEST FOR REU SUPPLEMENT

The TANGO project involves a number of tasks that require sound judgment and understanding of complex project objectives, but no advanced preparation in mathematics (beyond discrete structures) or in computer science (beyond a reasonably solid ability to program). Assigning these tasks to qualified undergraduates would allow the PIs and the graduate students to concentrate on their focused research agenda: development of table-interpretation techniques, automated table conceptualization, conceptualized schema integration, and information-extraction algorithms—all in an effort to semi-automate the generation of ontologies.

Student Involvement

For this REU supplement, we are requesting funds for two students.

Student #1. Data Reconciliation Project

Given that we will have achieved our research objective of precisely identifying which sets of data are to be integrated, we are left with the task of reconciling the representation of the data values. In the spirit of always being able to manually complete any task, *Student* #1 would build a simple-to-use interface that would provide for the following features.

- 1. Display of string values extracted from a table for a chosen lexical object set in the growing ontology.
- 2. Display of multiple sets of extracted string values for lexical object sets identified as "to be merged" in the growing ontology.
- 3. Access to a library of standard value canonicalization routines—translation programs that convert external string representations to internal representations. This access would include the ability to write code for a new canonicalization routine and add it to the library, and to select a canonicalization routine for a lexical object set.
- 4. Access to a library of standard output conversion routines—translation programs to convert internal representations to strings for output display. Likewise, this access would include the ability to write code for a new output-conversion routine and add it to the library, and to select an output-conversion routine for a lexical object set.
- 5. View and use of a side-by-side comparison interface for object-identity reconciliation.
- 6. View and use of a mechanism for value resolution for functional constraints.

As a step beyond manual intervention for data reconciliation, we would want to build in default actions and suggested resolutions. Default actions would allow for full automation (at the expense of possibly not reconciling data as an expert would), and suggested resolutions would allow for semi-automatic user intervention to direct data reconciliation (synergistically, with as much of the process being automated as possible).

Doing this work will help *Student* #1 to (1) learn about the problems involved in data reconciliation, (2) see the possibilities for automating data reconciliation, and (3) be involved with PIs and graduate students in their efforts to realize the full potential of TANGO—a grand research objective worth pursuing.

Student #2. Project Component Integration

The graduate students on the TANGO project have been doing research on several major, independent components. Except for agreeing on interfaces for the exchange of data, however, they have not focused on making the entire project work together. For TANGO to work properly, not only must each component work, but also the flow among components must work, resulting in the overall generation of an ontology. In addition, to the basic flow within the TANGO project, which results in a generated and growing ontology, several auxiliary components can also be incorporated into a larger project for semi-structured data management. These auxiliary components include alternative ways of generating integratable mini-ontologies (e.g., from sibling tables in hidden web pages and from user-specified forms); graph-layout software for managing growing ontologies; translation of our proprietary ontology representation to standard ontology representations (e.g., OWL); tools to allow users to present data subsets in table views and form views; links to original sources for data provenance; generation of extraction ontologies for data retrieval from unstructured and semi-structured web pages; generation of semantic mark-up for creating semantic web pages; and query processing software (both with standard query languages and a free-form, ontology-based query system, as well as table-based and form-based query systems). These are all projects we have worked on or are currently working on in our research group. We have not, however, been able to integrate them together into a workbench of tools for semi-structured data management.

To integrate these separate projects into a unified whole, Student #2 would need to creatively do the following.

- 1. Develop an overall framework for the TANGO project.
- 2. Add each major component of TANGO, allowing each component to operate manually, semiautomatically, or fully automatically.
- 3. Design an interface to allow a user to see the creation of an ontology as a unified whole.
- 4. Add components for experimentation with subjects (e.g., keystroke tracking software, event timestamps, results-evaluation views for experimenters).
- 5. Integrate auxiliary components into a larger workbench for managing semi-structured data.
- 6. Trouble-shoot the system—there will always be bugs to fix and problems to resolve.

This project involves some fascinating work for *Student* #2: (1) realizing the overall dream for TANGO; (2) taking research results from disparate projects and making them work together harmoniously; (3) opening up possibilities for larger semi-structured data management systems; and (4) as part of the integration, having the opportunity to add insights that are more than just the sum of the parts—as components come together, there will surely be opportunities for synergism among the individual parts that could not have been realized within their own boundaries.

Undergraduate Mentoring Involvement

Brigham Young University attracts excellent undergraduates; its freshman profile for 2006 includes 90% with ACT scores between 24 and 30 and an average high school GPA of 3.76/4.0. BYU is ranked 10th in the nation in the number of graduates who go on to earn doctoral degrees. Part of the reason for this excellent record is its aggressive support for undergraduate mentoring:

An extremely important aspect of research at BYU is the institution-wide emphasis on mentoring undergraduates involved in research projects. Undergraduates assist faculty with research and creative work on campus, often working in labs and co-authoring papers with faculty members. The students' work, which often results in presentations, performances and publications in academic journals, provides them with key credentials for graduate school admission or job placement. BYU undergraduates routinely win awards for their research while competing against graduate students.

In addition to BYU's excellent record with undergraduate research assistants, the PIs of the TANGO project have also been personally involved in mentoring undergraduates. Co-PI Lonsdale has had the opportunity to work extensively with undergraduate students since his arrival at BYU in 1998. He directs three research groups that involve mostly undergraduate participants, and his groups have been awarded four competitive mentoring environment grants from the university and college. He has published five peer-reviewed publications with undergraduate student co-authors. His collaborations have led to 22 conference presentations (9 at international venues) with undergraduate student co-authors (24 different students in total). Lonsdale has also chaired or refereed four undergraduate honors theses. He has also served as mentor for at least a dozen competitive undergraduate individual research grants funded by the university. Although PI Embley has mostly mentored graduate students, he has had and currently has undergraduate research assistants in his lab.

The undergraduate students will be full participants in our research alongside our graduate students. They will join our project meetings, make written and oral reports, and have weekly individual conferences with the PIs. They will also be provided with equipment and assigned space in our research lab.

Process and Criteria for Selecting Students

Embley and Lonsdale have been successful in recruiting, mentoring, and graduating female students. Currently, in our research group, three of our seven graduate students are female (a percentage much higher than the average in our computer science department). In addition, three female graduate students have completed graduate degrees under our tutelage in recent years (again, a much higher than usual percentage). We are not likely to be able to recruit from other under-represented groups—the percentages of other groups at BYU is small.

One undergraduate student (Jeff Peters) is currently working in our research lab. With additional support (through this REU supplement), he may choose to continue working with us. His resume is in the Supplementary Documents section of this proposal.

Specific recruiting efforts for either one or two other undergraduate students will include talking to students in our undergraduate classes, posting notices on bulletin boards, and sending email messages to targeted groups.

	ET Y	′E <u>AR</u>	3			
ORGANIZATION		POSAL	NO.	DURATIC	JN (months)	
Brignam Young University					Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR		A	WARD N	0.		
DAVID W EIIIJIEY		_NSF Fund	ed		Funds	Funds
(List each separately with title, A.7, show number in brackets)	CAL	Person-mo	nths CLIMP	Req	uested By	granted by NSF
1 David W Embloy, nono				¢	000361	¢
2. Davia V Ellipity - Iloite	0.00		0.00	φ	U	φ
	0.00	0.00	0.00		U	
3.						
5						
6. (1) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00		0	
7 (2) TOTAL SENIOR DERSONNEL (1 - 6)	0.00		0.00		U 0	
P. OTHED DEDSONNEL (SHOW NI IMPEDS IN PDACKETS)	0.00	0.00	0.00		U	
	0.00	0.00	0.00		0	
1. (U) POST DUCTORAL SCHULARS	0.00	0.00	0.00		0	
2. (U) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00			
3. (U) GRADUATE STUDENTS					<u>U</u>	
4. (U) UNDERGRADUATE STUDENTS					0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						
6. (0) OTHER					0	-
TOTAL SALARIES AND WAGES (A + B)					0	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					0	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					0	
					0	
					U 0	
2 FOREIGN					0	
F. PARTICIPANT SUPPORT COSTS				-		
1. STIPENDS \$12,000						
2. TRAVEL 0						
3. SUBSISTENCE 0						
4 OTHERO						
G. OTHER DIRECT COSTS			-			
1. MATERIALS AND SUPPLIES					0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION	2 PUBLICATION COSTS/DOCUMENTATION/DISSEMINIATION					
3 CONSULTANT SERVICES					0	
					0	
					0	
					U	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
(Rate: , Base:)						
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)					12,000	
K. RESIDUAL FUNDS					0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				\$	12,000	\$
M. COST SHARING PROPOSED LEVEL \$ 0 AGREED LE	VEL IF	DIFFERE	NT \$			
PI/PD NAME			FOR	NSF US	SE ONLY	
David W Embley		INDIR		ST RAT	E VERIFI	
ORG. REP. NAME*	D	ate Checked	Dat	e Of Rat	e Sheet	Initials - ORG

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

	ст (Cu <u>mulat</u>	ive			<u></u>	
Brigham Voung University		JFUSAL	Proposed		d Granted		
PRINCIPAL INVESTIGATOR / PRO JECT DIRECTOR		0	Filiposed	J Granieu			
David W Embley				0.			
A SENIOR PERSONNEL: PL/PD Co-PL's Faculty and Other Senior Associates		NSF Fund	ed		I Funds	Funds	
(List each separately with title, A.7. show number in brackets)	CAL	ACAD		Req	uested By	granted by NSF (if different)	
1 David W Embley - none	0,12		0.00	\$	<u> </u>	\$	
2. Dervie Lonsdale	0.00		0.00	Ψ	0	•	
3.	0.00	0.00	0.00		0		
4.							
5							
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0 00	0.00		0		
7.(2) TOTAL SENIOR PERSONNEL (1 - 6)	0.00		0.00				
B OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)	0.00	0.00	0.00		J		
	0.00	0.00	0.00		0		
	0.00		0.00	0 0			
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00				
4. (U) UNDERGRADUATE STUDENTS					U	+	
5. (U) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					<u>U</u>		
					0		
					U		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)					<u>U</u>		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)					0		
TOTAL EQUIPMENT E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSE 2. FOREIGN F. PARTICIPANT SUPPORT COSTS 1. STIPENDS \$ 12,000	SSION	5)			000000000000000000000000000000000000000		
3 SUBSISTENCEO							
4. OTHER0							
TOTAL NUMBER OF PARTICIPANTS (1) TOTAL PAR	TICIPA	NT COST	S		12,000		
G. OTHER DIRECT COSTS	-				,		
1. MATERIALS AND SUPPLIES					0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION					0		
3. CONSULTANT SERVICES					0		
4 COMPLITER SERVICES					0		
5 SUBAWARDS					0		
6 OTHER					Ŭ		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					12,000		
					0		
					12,000		
				¢	10.000	¢	
				\$	12,000	\$	
VI. CUST SMAKING PROPOSED LEVEL \$ U AGREED LE		UIFFERE	INI \$	105.11			
	\vdash					047/01	
Davio w Emdiey							
ORG. REP. NAME^		ale Checked	Dat	e Or Rat	e Sneet	Initials - ORG	

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

All \$12,000 is to be used for undergraduate RA salaries. The activities to be carried out are those listed in our request for an REU supplement. BYU does not require any administrative overhead for REU's.

Jeff Peters

jeffpeters@byu.edu 180 N 600 W, Orem, UT 84057 Cell: 801-885-9408 • Home: 801-226-2699

Experience

- PI- Engineering , *Software Developer: Summer internship* June 2007- Aug 2007 Orem, UT (Brad Daw: 1-801-735-7425)
- BYU- Comp-Sci Dept, *Research Assistant: TANGO Project* Jan 2007- Current TMCB, Provo, UT 84602 (Dr. Embley: 1-801-422-6470)
- BYU- Comp-Sci Dept, *Teaching Assistant: CS236* Sep 2006- Dec 2006 TMCB, Provo, UT 84602 (Dr. Embley: 1-801-422-6470)
- Bridge Technologies, *Driver / Laptop Repair Technician* Aug 2006– Sep 2006 480 North Freedom Blvd, Provo, UT 84601 (Chris Edwards: 1-801-377-6557)
- Hospitality Marketing Concepts, *Telemarketing* May 2004 377 East 800 South, Orem, UT 84058 (1-800-377-8921)
- Western Wats Research Center, *Phone Surveys* July 2002 March 2003 208 North Orem Blvd, Orem, UT 84057 (801-235-7084) *Employee of the Month* October 2002

Achievements

- Served honorable full-time mission in Eugene Oregon (LDS church), July 2004-2006
- Eagle Scout, with one Bronze Palm
- CompTIA A+ Certified Professional, for computer hardware and OS technology
- CompTIA Network+ Certified Professional
- Certified Novell Administrator (CNA) for Netware 5.1 Administration
- Cisco Certified Network Associate (CCNA) 2003, expired
- Invited to the 2003 Future Business Leaders of America National Leadership Conference in Dallas, Texas. *Placed 4th nationally in Networking Concepts competition*.

Education

- BYU Student Summer 03-Winter 04, Fall 06-Current, Expected Graduation: Winter 09
- Undergraduate Student in Computer Science with Major GPA: 3.38 Overall GPA: 3.26
- Mountain View High School Graduate of 2003 with GPA: 3.87
- Mountainland Advanced Technology College 2001-2002, 2002-2003 School Years

Abilities

- I have experience programming in Java (4 years), C++ (4 years), Pascal (2 years), Visual Studio .NET (1 year) and helping others find and fix bugs in their code
- Experienced in troubleshooting and fixing computer hardware and software and network problems on school, neighborhood and family computers for past 9 years
- Used source code control tool such as Subversion, and code coverage tool Cobertura

Areas of Interest

• Computer Repair, Programming, Networks, Math, Scouting, Service