Interpreting and integrating mismatched data on-the-fly during emergency response situations

Alan Bundy Fiona McNeill

I. Previous Track Record

The Mathematical Reasoning Group. Bundy and McNeill are members of the Mathematical Reasoning Group (MRG)[1], which is part of the Centre for Intelligent Systems and their Applications (CISA) within the School of Informatics at the University of Edinburgh. In the 2008 RAE, the School produced more world-leading (4*) research than any other computing department (69% more than the leading competitor), contributing 10% of the UK's world-leading research.

Since the mid 1970s, the MRG has been engaged on the computational analysis, development and application of mathematical reasoning processes and their interactions. Its work is characterised by its unique blend of computational theory with artificial intelligence. The processes studied by MRG have included theorem proving via proof methods, proof patching, analogy, symmetry, abstraction, diagrams and reflection; the learning of new proof methods, the formalisation of informally stated problems, the formation of concepts and conjectures and the interaction of automated systems with human users. The applications of MRG's work have been to: proof by mathematical induction and co-induction; analysis, including non-standard analysis; mechanics problems; the building of ecological models; the synthesis, verification, transformation and editing of both hardware and software, including logic, functional and iterative programs and process algebras; the configuration of hardware; game playing; and cognitive modelling. More recently, the group has also focussed on representation change and ontology evolution, developing novel work on dynamic, structural change. This has been applied in several domains, such as multi-agent communication and the Semantic Web, ontology evolution in legal reasoning, theory evolution in physics and automated cleaning of automatically generated ontologies.

The group has been supported by over 50 research grants, including an EPSRC rolling funding grant (1982-2002), and three EPSRC platform grants (2002-2007; 2007-2011; 2011-2015).

Professor Alan Bundy: has been active in mathematical reasoning research since 1971 and has become a world authority. His international reputation is witnessed by his being made a founding fellow of both of the two international AI societies: AAAI and ECCAI, in addition to the UK society, AISB, and serving terms as Chair of both IJCAI Inc and CADE Inc. He is also a Fellow of the Royal Society of Edinburgh, the Royal Academy of Engineers, the British Computer Society and the Institution of Engineering and Technology. He won the SPL Insight Award in 1986, the IJCAI Distinguished Service Award in 2003, the IJCAI Research Excellence Award in 2007, the CADE Herbrand Award in 2007, was an SERC Senior Fellow (1987-92), a member of the Hewlett-Packard Research Board (1989-91), Head of the Division of Informatics at Edinburgh (1998-2001), a member of the ITEC Foresight Panel (1994-96), a member of both the 2001 and 2008 Computer Science RAE panels (1999-2001) and (2005-2008), a member of the Scottish Science Advisory Council (2008-date) and was the founding Convener of UKCRC (2000-2005). He is a Vice President and Trustee of the BCS, with responsibility for the Academy of Computing, the UK's learned society for computing. He is the author of over 230 publications and has held over 50 research grants.

Dr Fiona McNeill: is a research associate and developer of the system ORS (Ontology Repair System), which performs on-the-fly structural matching during agent interaction. She has worked on the OpenKnowledge project, focussing on matching and on e-response simulation in a flooding disaster, and is now work on an ONR-funded project developing sound mathematical underpinnings for Ontology Match-

[1] http://dream.dai.ed.ac.uk/

ing. She was the programme chair of the Workshop of Matching and Meaning (WMM09 and WMM10) and co-chair of the Workshop on Discovering Meaning on the Go in Large Heterogeneous Data (LHD-11) at IJCAI11, and is on several programme committees (including OM'09, OM'10, OM'11).

Suitability of the team

McNeill and Bundy have worked together on developing dynamic matching techniques since 2001, first during McNeill's PhD, in which the ORS system was first developed, then through the EU-STReP-funded OpenKnowledge (OK) project, where McNeill applied the ideas developed in the PhD to the domain of web service integration, with Bundy acting in a supervisory role, to their current ONR-funded grant which focusses on developing ORS further and working on the theoretical underpinnings of dynamic matching. They have also supervised (separately and jointly) many PhD, MSc and UG students developing this work.

McNeill has experience in the emergency response domain, being the lead developer of the emergency response simulator, the main testbed in the OK project, and developing research focussing on the information sharing needs during a flooding scenario in that context. We are also greatly assisted by our collaboration with the Artificial Intelligence Applications Institute (AIAI), a technology transfer organisation within CISA which is the home of the Planning and Activity Management group, which has a strong focus on planning in crisis and emergency situations. We have worked closely with them in the past and will be able to consult them frequently and easily.

We are also very fortunate to have access to significant local expertise in data intensive systems. The National e-Science Centre is located in Edinburgh, and we have presented our work to them previously. The Data Intensive Research group, lead by Professor Malcolm Atkinson, is part of CISA and is also part of the National e-Science Centre. We have previous experience of working with members of this group, and envisage collaborative work with them during this project.

II. Case for Support

2.1 Introduction

We exist in a world of large data: most organisations have large data stores, and many (such as governments) have enormous data stores. Accessing and utilising this data quickly and effectively is essential for many real-world tasks. One area where these needs are particularly pressing, and which is of particular concern for the military and for the nation in general, is in the domain of emergency response. The military is becoming increasingly involved in such events, and "major natural hazards and accidents" were identified as one of the four Tier 1 risks in the 2010 National Security Strategy. The urgent and unpredictable nature of these events mean that they are hard to prepare for, and that responders may find themselves needing to interact with a diverse set of different organisations. The difficulties of sharing information under such circumstances has been widely recognised, for example in the Pitt report into the 2007 floods [7], and the government response to it [2].

In such an emergency response scenario, many agencies - some of which may have extensive experience of working together and some of which will not have interacted before - need to quickly and efficiently exchange knowledge in order to provide an effective response. One of the great difficulties of such automated knowledge sharing is that each participant will have developed and evolved its knowledge sources independently and there will be significant variation in how they have done this. These differences may be of various types, for example:

- different agencies may use different words for similar things: e.g., `agency1` uses the term *person* where `agency2` uses the term *individual*;
- different agencies structure their data differently: e.g., `agency1` uses the term *person(location,age)*, considering a person to have two relevant attributes: *location* and *age*, whereas `agency2` may use the term *individual(age,location,vulnerability)*, considering a person to have three relevant attributes: *age*, *location* and *vulnerability*.
- The specificity of words may differ: e.g., one agency may distinguish *civilians*, *combatants*, *personnel*, etc., where another agency represents only *person*.
- The organisation of terms may differ: for example, one agency may have *person(age,location,vulnerability)*, where another may have *person(name,person_id)*, *location(person_id,person_location)*, *age(person_id,years)*, *vulnerability(person_id,person_vulnerability)*.

When these different agencies interact, their data is not mutually comprehensible. For example, if `agency1` needs data about all the people in the disaster area, it may send out a request *person(location,age)* to other agencies involved in the response, to gather what information they have about affected people. But `agency2` would not be able to automatically interpret this query because it will have nothing in its database that directly corresponds to it. A human user may be able to search through `agency2`'s data to try to find a match, but this will be extremely time-consuming and difficult, because databases are hard for humans to read, and because data sources may be very large. In an emergency response scenario, finding such information quickly and reliably is essential. It is therefore necessary to find ways of doing this automatically. We propose to adapt techniques designed for dynamic ontology repair to address this problem. The hypothesis of this project is:

> *The use of ontology repair techniques to interpret mismatched information during a disaster-response situation allows data from large and heterogeneous sources to be shared quickly and efficiently.*

2.2 Dynamic structural and semantic matching

We propose to develop techniques to perform approximate matching to interpret data from one source in terms of another on-the-fly in real time. When there is no exact match in a user's data source for a required term, such as *person(location,age)*, our techniques will discover whether there is a term in that data source which is approximately equal to this term, and, if so, how it differs. This approach will use a combination of semantic and structural matching techniques to identify, for example, that *individual(age,location,vulnerability)* could be mapped to the query term to provide the required information. This potential match could be provided to a human user attached to `agency2` for confirmation before the requested data is provided - if the agency decided that the required level of accuracy and the time available made it desirable to have human approval.

Whilst there has been considerable work in the field of Ontology Matching (see, for example, [3]), this has been largely devoted to the problem of finding full maps between large ontologies, and is performed offline, often with a large manual component. It therefore does not address the problem at hand, where matching needs to be focussed (that is, a match needs to be found in the database for a specific query, rather than matching an entire knowledge source to the database), dynamic (that is, it needs to be done quickly, during interaction) and automatic (because of time constraints, and also because having human users with the required level of skill available at all times by all agencies would be infeasible).

Our approach allows us to deal with the problem of matching large data sources in real time by ignoring the vast majority of that data. Matching need only be performed *on demand*, when a particular term needs to be interpreted, and the information contained in the term allows us to narrow the search considerably (for example, only terms with predicate names which match or are related to the one contained in the search term need to be considered). Therefore, the problem of matching two or more large data sources is reduced to the problem of matching a single term against a small subset of each of these data sources. This enables the problem to be addressed in real time. Because of these time restrictions, and because of the costs associated with human involvement, it is desirable that this process is fully automated. However, matching is an inherently difficult problem because meaning is slippery: it is often not possibly to say, out of context, whether one thing means exactly the same - or more or less the same - or not at all the same - as another thing. In some domains, the lack of certainty may be acceptable: information, that will mostly be correct, will be shared quickly and efficiently. However, in some domains there is a high cost to information being incorrectly or inappropriately matched. Thus, in such domains, whilst the matching should be fully automated, the proposed matches should not be put into use until they have been approved by expert users. In such cases, it will be necessary to present the potential matches to users in a highly intuitive way, explaining the implications of the match (for example, which words are being matched, what information may be lost, etc.) and allowing users to approve of potential matches, and sort through different possibilities.

We have extensive experience of the problem of focussed, dynamic ontology matching (see, e.g. [5, 6]), and have developed a successful system, ORS (see, e.g., [4]), to deal with this problem in the context of agent interaction. The potential for the techniques on which this work is based to be utilised as a front-end for exploring large data, and the benefits of doing this, have been increasingly clear[2]. The domain of emergency response provides an excellent testbed for these ideas, because the need for such techniques is very pressing and many end-users are keen to collaborate. We have developed collaborators within Dstl[3], the Met Office[4], Strathclyde Fire & Rescue[5] and Scottish Resilience[6], who are keen to assist this work and provide feedback and data for us.

The aim of the proposed project is to investigate these problems and build a proof-of-concept system to determine how applicable our ontology matching techniques are within this domain. We will use real data from our end-user collaborators, together with feedback about real problems encountered on-the-ground both to identify what sort of problems need to be addressed and to evaluate our system once completed.

Note that this project is intended to be ex-

[2]This was explored, for example, in the recent IJCAI workshop *Determining Meaning on-the-go in Large Heterogeneous Data* (LHD-11) (http://dream.inf.ed.ac.uk/events/lhd-11/), co-chaired by McNeill and sponsored by ONRG, W3C and Yahoo!Research. It included a panel session on "Big Society meets Big Data: Industry and Government Applications of Mapping Meaning" in which Tom McCutcheon of Dstl, together with John Callahan of ONRG, Peter Mika of Yahoo!Research and Thomas Steiner of Google, participated

[3]Contact: Tom McCutcheon *tgmccutcheon@dstl.gov.uk*

[4]Contact: Mark Gibbs *mark.gibbs@metoffice.gov.uk*

[5]Contact: Des Kingston *des.kingston@strathclydefire.org*

[6]Contact: Kerry Jardine *kerry.jardine@scotland.gsi.gov.uk*

ploratory. A complete solution to this problem, and a thorough analysis of the available data, would require a far larger project. The intended outcome of the project is an indication as to whether or not the hypothesis seems likely, rather than a full evaluation of the hypothesis. If the indication is positive, a follow-on project would be desirable which would contain a much more thorough examination of potential mismatches in relevant data and would develop a much more sophisticated system, dealing with more potential mismatches. A negative indication may still give us useful scope for future work, because an analysis of why the proof-of-concept system failed would lead us to a better understanding of the problem and potential solutions to it.

2.2.1 Evaluation

The evaluation of our approach will form a central part of the project. We will develop a plausible scenario, guided by our emergency response, and particularly our contacts at Dstl, in which different agencies are interacting using real data sets of a suitable size (i.e., as large as possible). Below we outline three sub-hypotheses which will enable us to judge whether our central hypothesis, described in Section 2.1, has been justified.

Evaluating the efficiency of the approach
Hypothesis: *Employing our matching techniques in plausible scenarios enables matches to be found from large data sets in real time.* We will determine the time it takes to find, present and approve appropriate matches during the execution of our scenario, and consider whether this can be considered to be 'real time'.

Evaluating the precision of the approach
Hypothesis: *The matches produced by our system will generally be of high enough quality to be useful, or if our system fails to find a suitable match, this is because no such match exists.* This is not simply a question of determining whether or not the system does what it is supposed to do: even if the system does return the matches we intend it to, it is still essential to determine how useful such matches really are. It is only possible to understand this once we see what implications these matches would have in a realistic scenario. Whilst a poor outcome for the first evaluation criterion may suggest that we need to improve our system, a poor outcome here may raise more fundamental questions about the extent to which context-free automatic matching is possible, and may provide valuable insights into how this problem should be approached.

Evaluating the usability of the approach
Hypothesis: *It is possible to present potential matches to expert users in such a way that they can quickly and easily understand the implications of these matches and approve appropriate ones.* This is primarily a question of usability: can we present potential matches to expert users in a way that allows them to quickly and easily understand the implications of approving or choosing such matches, without interrupting the flow of the scenario excessively? This part of the evaluation will require contact with such expert users, who will come from Dstl and our emergency-response contacts. However, to maximise the usability of our approach would require considerable effort on the interface design, which is outside the scope of the project. The feedback we receive from this will therefore be useful to highlight potential problems and to analyse whether enough information is available for experts to make appropriate decisions, but is likely to significantly underestimate the full potential of the system.

2.3 Programme of Research

We describe the proposed plan of work in terms of the work-packages, their goals, timescales (in person months) and deliverables. The scheduling and significant dependencies are shown in the diagrammatic workplan.

Work Package 1: Analysis of Relevant Data Sources We will work with our collaborators to identify data sources which are accessible and have been used in such contexts, and to develop a picture of the kind of data sharing problems they have encountered[7]. This will allow us to develop a small set of real mismatch problems with which to develop our techniques. An initial attempt to gauge how frequently such mismatches occur will be made, though a thorough statistical analysis of the data is outside the scope of the current project.
Time: McNeill: 1.5 months, Bundy: 1 week
Deliverables: A set of mismatch problems. A workshop paper describing the results of our in-

[7]Our interactions thus far with our collaborators suggests that such examples will not be hard to find.

vestigation.
Dependencies: None **Work Package 2: Adaptation of repair techniques** Whilst we are developing the problem set described above, we will be considering how our existing repair techniques could be applied to them, and what level of adaptation they will require. We expect many of the existing techniques to be relevant, because they are very general purpose (for example, considering name changes for predicates or arguments, or changes in the arity of predicates). We will also collect a set of mismatches that are not addressable by our current techniques, if they exist. We will not create new techniques to address them in this current project, but this would be an obvious basis for extending our work.
Time: McNeill: 0.5 months, Bundy: 0.5 weeks
Deliverables: A set of adapted repair techniques, which are likely to be very similar to our existing techniques, together with a set of unaddressed mismatches.
Dependencies: WP1 (which must be begun, though need not be completed) **Work Package 3: Development of proof-of-concept system** We will develop a proof-of-concept system to investigate the application of these techniques to our data. The design will be informed by our experience implementing these techniques in a different context, but will require alteration to the mismatch diagnostic technique, which was designed in a multi-agent environment. We will also design a simple interface for the user (e.g., command line prompts), and determine how to provide appropriate information for this user to allow them to make relevant decisions.. **Time:** McNeill: 2.5 months, Bundy: 3 weeks
Deliverables: A proof-of-concept system. A conference paper describing it.
Dependencies: WP1, WP2 **Work Package 4: Evaluation of system** Once the system has been completed and tested, it will be evaluated against a scenario devised from existing data supplied by end-user collaborators. Again, this evaluation will be exploratory and will be intended to give an indication of the usefulness and applicability of the ideas.
Time: McNeill: 1.5 months, Bundy: 1 week
Deliverables: A journal paper describing the completed project, including the evaluation. A document, possibly in the form of a proposal, exploring what the project has revealed about what the next steps should be and how this work should be expanded into a more complete solution to the problem.
Dependencies: WP3

2.4 Project Management

The project will be directed and managed by Professor Bundy. He will devote 10% of his time to working in a supervisory capacity on this project. Dr McNeill, will be employed to work on this project at 50% time. Thus the (approx) 7.5 months person time allocated in the work-packages described in Section 2.3 will take place over a 12 month period, with McNeill providing 80% of the effort, and Bundy 20%. Bundy and McNeill have a long history of successful interaction, and will meet frequently during the project.

The key milestones and deliverables of the project are set out above and have been designed to ensure that they are specific, measurable and attainable within the timeframe of the project. Frequent monitoring will occur to ensure that these will be met. Some of these (for example, the journal paper), will not be completed until after the end of the project; however, all the implementation and evaluation will be completed within the allotted 12 months.

Since we are tackling an aspect of a very large problem within a fairly short project, it is essential that the scope of the project is clearly defined and strictly adhered to. This will include:

- Initially observing a large number of datasets in which we might find potential mismatches, but quickly reducing these to a small number of datasets which will be more carefully analysed;
- Restricting the implementation (but not the analysis) to mismatches that are similar to mismatches for which we already have well defined diagnostic and repair techniques. The key research question will be how to adapt these diagnostic techniques to the context of the current environment, where some of the feedback that is possible in the multi-agent environment they were originally designed for will not be available.
- Designing the user interaction with the system, to view and approve of potential matches, so that it is very simple and

based on command line prompts. Providing a high-quality interaction experience to the user is essential if this technology is to become useful in the field, but it is not essential to assess the potential of this approach, and is outside the scope of this project.

- Restricting the scope of the evaluation (for example, the size and number of the data sets on which we will perform the evaluation) to something which is performable within the time limit.

All these restrictions are likely to provide great possibilities and inspiration for further work. However, rather than allowing these extensions to creep into a short project, we will instead carefully catalogue them with a view to writing a proposal for a longer project if the results of this project are promising. We believe a project with this scope will be sufficient to investigate the potential benefits of the approach, and to give us a clear idea of what the next steps will be, as well as being achievable within 7.5 person months.

Interaction with end-users will be most important during work packages 1 and 4, and will consist of at least one face-to-face meetings with each collaborator during each of these work-packages, and additional contact by phone and email. Travel costs to cover these visits are included in the costings.

We believe that, due to the scope of the project, the risks associated with it are not high. One risk is that we may find it difficult to access a suitable number of data sources; however, the positive interactions we have had with our collaborators thus far indicates that this is unlikely, and the number of data sources we require in this project is not large. Another risk is that the kinds of mismatches we discover are not compatible with our matching techniques; however, the generality of our techniques makes this unlikely, and any mismatches we discover that our not covered will provide an excellent basis for further work. A third risk is that the evaluation indicates that the matches produced are too vague or approximate to be useful. This would be a disappointing outcome for our own work, though it may provide useful insights into how we could improve our work. However, from the point of view of understanding this problem more deeply and moving closer to a solution which might solve this problem, this would be a very valuable result, and one which should be of great use to other researchers.

2.5 Relevance to Beneficiaries

Military applications The evaluation of this project will focus primarily on an emergency response scenario, such as the military are frequently called on to lead or participate in. Moreover, the application of the techniques developed in this project extend to many other military applications, where fast and timely interaction between disparate agencies with incompatible information is crucial.

Living with environmental change The UK Government has recognised that we cannot prevent major damage and disruption from the impact of natural disasters such as flooding on national infrastructure. It is therefore critical rescue services, central government agencies and a host of other providers have as much support as possible to provide as much support as possible to prepare for and manage those adverse events that do occur. This project seeks to apply leading techniques from a theoretical area of computing science to a pressing strategic problem for UK society.

Non-academic collaborators

The key issue of this project is to apply abstract reasoning techniques to identify potential problems in the transfer of information that have been a characteristic of the UK response to previous major disaster events. Our work will seek to ensure that such problems do not mar our future response to emergency situations.

Relevance beyond the e-response domain

More generally, the UK government is in the midst of a programme of information modernisation through the provision of web service architectures. However, much of this work focuses on information push without sufficient regard to the consumers of that data. In consequence, information is often poorly structured in terms of the end-user requirements. Our use of matching techniques provides a generic technique for the use of table-top exercises and theoretical analysis to explore the relationships between government information producers and consumers that has a growing relevance for e-government beyond the emergency services.

2.6 National Importance

The problem of fast, efficient and secure data sharing in a world of large, heterogeneous data sources is one that is becoming increasingly pressing, and it is strongly in the national interest to deal with these problems. In particular:

- the cost to the country, socially and economically, of inefficient disaster response, is vast. Whilst it is not possible to nullify the impact of such events, enabling relevant responders to access and share relevant data without delays and confusion can mitigate against the worst effects of these events, and reports into the response to these events (such as the Pitt report and the government response to it) highlight that this is greatly needed but is currently far from happening.
- the ability to absorb and use new data from various sources is also crucial in military situations, and effective responses are of significant national importance.
- the economic and social impact of the development of the World Wide Web has been enormous and has greatly benefited the country. It is anticipated the development of an effective and usable web of data (the Semantic Web) could have almost as great an impact. Thus far, the myriad difficulties associated with data sharing mean this vision has not been realised. Being at the forefront of bringing about this new world is of key national importance.

We believe that our work could play a useful part in bringing about these objectives. The original Semantic Web vision of unique identifiers and shared ontologies has proved to be impractical, and the academic community is still attempting to deal with the implications of this. Whilst the majority of effort has focussed on making data interoperable prior to interaction (for example, in the Ontology Matching community), we have been at the forefront of research that focusses on the need to make data interoperable during interaction (for example, Bundy and McNeill presented these ideas in an invited paper in the Dartmouth Conference 50th anniversary commemorative journal aiming to identify the key directions for AI over the next 50 years [1]). There are clearly many important aspects to this problem of data sharing which this research does not address, but the problem that it does address - that of identifying when data from disparate sources is the same or similar even when it is expressed using different words and structure, and understanding what these differences imply - is one central aspect of the problem. We believe we are therefore in a unique position to improve the ability to share heterogeneous data on the fly, and thereby bring about these key national benefits.

2.7 Dissemination and Exploitation

The continued development of a community spanning academia, industry and government, and frequent interactions within this community via workshops, personal interactions, tutorials and publications, is central to our approach. The good understanding of our, and others, work which is facilitated by this community will greatly enhance the uptake of this work within, and beyond, this community. We will also ensure that our work is freely available by maintaining a website from which analysis, data[8], software and publications can be downloaded. Further details can be found in the Pathways to Impact document.

[8] where appropriate

References

[1] A. Bundy and F. McNeill. Representation as a fluent: An AI challenge for the next half century. *IEEE Intelligent Systems*, 21(3):85–87, 2006.

[2] The Government's response to Sir Michael Pitt's review of the summer 2007 floods. DEFRA, HMSO, London, 2009.

[3] Jérôme Euzenat and Pavel Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[4] F. McNeill and A. Bundy. Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *IJSWIS*, 3(3):1–35, 2007. Special issue on ontology matching.

[5] Fiona McNeill, Paolo Besana, Juan Pane, and Fausto Giunchiglia. Service integration through structure-preserving semantic matching. *Cases on Semantic Interoperability for Information Systems Integration*, 2009.

[6] Fiona McNeill and Alan Bundy. Facilitating virtual interaction through flexible representation. *Encyclopedia of E-Business Development and Management in the Global Economy*, 2010.

[7] The Pitt review: Lessons learned from the 2007 floods. HMSO, London, 2007.