Working Notes: Multi-Agent Systems in Browser Automation

> ***Preliminary thoughts on extending the Generic User framework***

Agent Types to Consider

Core Agents

- **Domain Specialist**: Application-specific knowledge and goals
- **Browser Interaction Specialist**: UI navigation and interaction expertise

Potential Auxiliary Agents

- **Memory Agent**: Long-term storage of interaction patterns and context
- **Planning Agent**: High-level strategy for multi-step tasks
- **Error Recovery Agent**: Specialized in handling unexpected states
- **Security Agent**: Verifies safety of interactions and data handling
- **Learning Agent**: Identifies patterns for model improvement

Memory Agent Implementation Concepts

Function

- Listens to all agent conversations
- Stores significant information with temporal relevance
- Provides contextual recall without disrupting primary interactions
- Maintains "institutional memory" across browsing sessions

Storage Architecture

- Short-term working memory (current session)
- Medium-term memory (recent sessions)
- Long-term memory (persistent patterns and important events)
- Forgetting mechanism with importance-based retention

Interaction Patterns

- **Side-channel notifications**: Providing context without interrupting main flow
- **Query-based access**: Agents explicitly request relevant memories
- **Proactive suggestions**: Memory agent interjects when confidence of relevance is high

- **Background enrichment**: Automatically enhances commands with historical context

Coordination Mechanisms

Token-Based Coordination

- Single "action token" passed between agents
- Only the agent holding the token can execute browser actions
- Prevents simultaneous conflicting operations
- Can create bottlenecks in complex scenarios

Hierarchical Supervision

- Supervisor agent oversees interactions between specialized agents
- Resolves conflicts based on domain expertise and priority
- Maintains overall goal coherence
- May create single point of failure

Market-Based Allocation

- Agents "bid" for action opportunities based on confidence and expertise
- Resources allocated to highest-value actions
- Potentially more flexible than rigid hierarchies
- Requires careful balancing to prevent manipulation

Federated Decision Making

- Distributed consensus mechanisms
- Voting or weighted opinion aggregation
- More robust to individual agent failures
- Higher communication overhead

Domain Boundary Management

Potential Approaches

- **Static domain allocation**: Pre-defined responsibilities
- **Dynamic negotiation**: Agents agree on boundaries during operation
- **Expertise-based routing**: Tasks routed to most qualified agent

- **Context-sensitive handoff**: Boundaries shift based on situation

Handoff Protocols

- Explicit success criteria for task completion
- Standardized state representation during transfers
- Acknowledgment mechanisms for successful handoffs
- Rollback procedures for failed transfers

Open Questions for Further Research

1. How does system performance scale with increasing number of agents?
2. What communication bandwidth is required for effective coordination?
3. How to prevent "agent collusion" that might work against user interests?
4. Can emergent behaviors develop that weren't explicitly programmed?
5. How does the system handle partial agent failures?
6. What debugging and observability tools would be needed for multi-agent systems?
7. How to measure and optimize the trade-off between coordination overhead and improved capabilities?
8. What are the implications for training data collection in a multi-agent environment?

Potential Experimental Approaches

- Start with simple two-agent system and incrementally add complexity
- Run A/B tests comparing different coordination mechanisms
- Create adversarial testing scenarios to identify boundary issues
- Develop metrics for coordination efficiency and conflict frequency
- Build simulation environment to test scaling properties before deployment

Related Work to Explore

- Multi-agent systems in robotics
- Organizational behavior theories
- Distributed systems coordination algorithms
- Cognitive science models of human team coordination
- Game theory approaches to resource allocation
- Biological models (ant colonies, immune systems)